THE GROWTH OF MATHEMATICAL IDEAS

Grades K – 12

Twenty-Fourth Yearbook

THE NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

WASHINGTON, D. C., 1959

Correspondence relating to and orders for
additional copies of the Twenty-Fourth
Yearbook and earlier yearbooks should be
addressed to

THE NATIONAL COUNCIL OF TEACHERS
OF MATHEMATICS
1201 SIXTEENTH STREET, N. W., WASHINGTON 6, D. C.

Manufactured in the United States

Preface

THIS YEARBOOK is intended to be primarily a highlighting of the essential elements of those basic mathematical understandings which should be continually developed and extended throughout the entire mathematics curriculum, grades K–12 and beyond. This is the content of Chapters 2 through 8.

Growth in understanding of a fundamental idea and of its extensions is a continuous process that is facilitated by properly chosen classroom techniques. Hence our second objective is to define and illustrate some classroom procedures and their psychological bases, which we feel are also appropriate and important at all levels of instruction. This you will find in Chapter 10.

In Chapter 9 we have tried to illustrate, if not to define, a major objective of mathematical training; namely, training in *mathematical modes of thought.* These modes are not quite mathematical concepts themselves, but are rather understandings and procedures which are implicit in the study of all mathematical topics.

They include ideas and processes which might by some be labeled *problem solving* and by others, *creative thinking.* Little if any of this can be taught as a separate topic in a text or a unit in a curriculum. Students can be helped to develop creativity and problem solving ability, we believe, only if their teachers constantly and repeatedly lead them to and through *problem solving* and *discovery* processes and encourage them to strike out mentally for themselves into problem situations or into areas new to them. Chapter 9, then, forms a bridge between the mathematical concepts of Chapters 2 through 8 and the methodological ideas of Chapter 10. Some of our most desirable objectives, while requiring an understanding of mathematical ideas by both teachers and pupils, will be achieved best by teachers who consciously adopt classroom procedures which develop mathematical thinking habits and techniques.

Our third objective is to assist teachers and supervisors to extend and apply the ideas of the book in their own situations. This is the objective of Chapter 11. A single book can only present a broad outline of the mathematics program in grades K through 12. Even if more details could be included in this book, they might, without careful study and interpretation, confuse rather than clarify that which is the real intent of this book. Included in Chapter 11 are some suggestions for extending the insights and testing the interpretations which a reader might be expected to get from this book.

It is inevitable that this twenty-fourth yearbook should be compared with and related to the twenty-third yearbook, *Insights Into Modern*

Mathematics. Although the twenty-fourth was neither conceived nor written as a supplement to the twenty-third, they should supplement one another well. We have sought to make this book a self contained unit insofar as any single book on such a broad topic can be. To do this we have written chapters entitled "Number and Operation," "Relations and Functions," "Proof," and "Probability" which in title overlap those of the twenty-third yearbook. However, where that book stressed mathematical theory and *modern* mathematics, we here are trying to suggest how basic and sound mathematical ideas (whether *modern* or not) can be made continuing themes that are recognized by teachers and pupils as a part of a spiraling growth and development of mathematical understandings from the kindergarten through grade 12 and beyond.

This yearbook was originally proposed to the Board of Directors of the National Council of Teachers of Mathematics, at its April 1954 meeting, by its Yearbook Planning Committee, F. Lynwood Wren, Francis G. Lankford, Jr., and Daniel W. Snader, as a yearbook on "in-service education for mathematics teachers." The committee's suggestion was adopted after the board modified its description by adding: "This book is not intended to be a textbook on methods, nor a survey of curricula, but rather a highlighting of the most basic mathematical themes which should be central to the entirety of a modern mathematics curriculum, and of the similarly key concepts of modern teaching techniques. These themes should be displayed as concretely as possible and at as many different levels of instruction as may be." Phillip S. Jones was then selected to be the editor of the book. The undersigned committee developed the framework of broad outlines and policies within which the authors have written.

The committee soon realized that to write concretely at a wide range of grade levels would take several authors with varied experience for nearly every chapter. It further realized that the book should itself have some unity and continuity which would require face-to-face discussion and even argumentation among its authors. Hence the committee suggested to President Marie Wilcox and the Board of Directors that funds be sought to pay for meetings of the writers and committee where the book's purposes, content, and organization could be jointly discussed, where illustrations and criticisms of first and second drafts could be shared, and where some actual writing could be done.

The board endorsed this proposal and urged President Wilcox to ask the National Science Foundation for funds. Funds for these purposes were granted in October of 1956. Whatever good this book may have would have been less without the discussions and revisions made possible by the funds allocated by the Foundation.

In addition to paying for such meetings, these funds also made it possible to reproduce over a hundred copies of an early draft of the book, to be read and criticized by a number of teachers in classes and seminars all over the country in the spring and summer of 1957.

We regret that some who helped us cannot be listed because some critiques were summaries of the thinking of groups and others were mailed to us without names; nevertheless, we wish to acknowledge with thanks the help of the following persons, who cannot, however, be held responsible for any of our errors or failures.

George E. Adams, Jr., Marjorie Y. Adams, Jackson B. Adkins, R. Ames, John B. Anglin, Elsie Baehre, Jessie Ayres Bailey, John Bartelt, James S. Beadle, Milton W. Beckmann, Roy O. Birmingham, Jr., Josephine Bodever, Alpheus Booker, Mary T. Boulware, Boyce Brown, John G. Bryan, Joyce Burchinal, Harold R. Burke, Barbara S. Burns, Richard B. Burns, Miriam Burris, Frank C. Carpenter, Jack Carpenter, Audrey D. Carr, Helen Chaffee, Dorothy Clark, John R. Clark, Arthur Collard, Thomas O. Corlett, Robert G. Crook, Mervyn Dauenbaugh, Nathan L. Davis, Dale Davison, Aristides Demetrion, Geraldine Dolan, Laura K. Eads, Harriett E. Emery, Kathryn Evans, William H. Evans, Walter J. Ferdon, Louise Fitchett, Carroll E. Fogal, William A. Gager, Paul R. Gagneaux, Ferdinand Gezich, John V. Ghindia, Jonathan Gillingham, Moses L. Glenn, Edith M. Glidewell, Richard Gorham, Elnice Greene, Jack A. Gustafson, Ruth M. Hannum, Irell S. Harp, Thomas Y. Harp, Arthur W. Harris, M. L. Hartung, Billy F. Hobbs, Virginia Hoge, Lucille Houston, Roy E. Howell, Jesse Humberd, Donovan A. Johnson, Robert L. Johnson, William L. Jonas, Robert Jorgenson, Clark Kaplan, Houston T. Karnes, Josephine Anne Kegerreis, Stanley Kegler, Kathleen Keller, Calven E. King, Helen Kriegsman, Donald E. Kuhnle, Mary Carolyn Labbe, Dorothy M. Leffler, Curtish Leicht, Catherine A. V. Lyons, Evan Maletsky, Ronald O. Massie, B. E. Meserve, T. L. Meyer, Dorothy R. Miller, Willis M. Miller, Howard Mion, Francis Mueller, T. Nelson, Frederick A. Parker, Philip Peak, Christine Poindexter, Frank P. Prather, H. Vernon Price, Louis E. Prieskorn, Mary Anne Prunier, Charles N. Race, Myrtle Rehwinkel, Robert L. Root, George H. Ross, Myron Rosskopf, Mary Lou Rohrbaugh, Velma I. Rust, Fred Schleiber, Marie T. Shires, Donald M. Silberger, Helen T. Simpson, James T. Sims, Sister Margaret, Joseph Sloboda, Jr., Janie Smallwood, Vernona L. Smukal, Roger F. Soucy, Andrew Stevenson, De Vere W. Stevenson, Alan D. Stewart, Edwin H. Stieben, Henry Swain, Dorothy Sylling, Walter Szetela, Julia Tegner, John E. True, Bruce R. Vogeli, Marguerite A. Watkins, Anna D. Weatherford, J. Fred Weaver, Catherine N. Wheat, Marie S. Wilcox, Edward J. Willis, and John Earl Wood.

Finally, all the writers and the committee, as well as Paul Clifford and Max Beberman, have shared in discussions of all chapters. However, since we occasionally failed to be in complete agreement, no writer nor committeeman can be held responsible for statements other than those in the chapter he wrote, but if there is any excellence in the book, credit for it must be shared with all who have worked so hard on it.

And, as the book goes to press, we feel we must add a word of sincere thanks to Myrl Ahrendt, Executive Secretary of the National Council of Teachers of Mathematics, who has labored hard and conscientiously to see our manuscript through the physical process which has made it into this book.

The Committee
Harold P. Fawcett
Alice M. Hach
Charlotte W. Junge
Henry W. Syer
Henry Van Engen
Phillip S. Jones, *Chairman*

Contents

1

The Growth and Development of Mathematical Ideas in Children

Or How and Why To Use This Book

PHILLIP S. JONES

We who are planning and writing this book have on occasion disagreed as to the relative importance of some of the goals of mathematics education, and as to the best methods for achieving them, but we are agreed that the following are axioms:

Axiom 1. *The best learning is that in which the learned facts, concepts, and processes are meaningful to and understood by the learner.*

Axiom 2. *Understanding and meaningfulness are rarely if ever "all or none" insights in either the sense of being achieved instantaneously or in the sense of embracing the whole of a concept and its implications at any one time.*

The sudden perception or flash of insight which is one of the joys of mathematics learning and teaching comes only to those who have, with thought, struggled to extend or apply concepts which have been partially understood earlier. Further, meanings and understandings themselves change continuously as they are extended, broadened, and applied in different situations. For example, the early elementary pupil who suddenly perceives that 9 plus any other single digit number is the teen number whose units digit is one less than the single digit number probably first sees this after having added a number of specific combinations such as $9 + 3 = 12$, $9 + 4 = 13$, $9 + 6 = 15$, and so on. Even if he had been told this rule, it is doubtful that it would have been really meaningful to him until after he had worked out several examples. Later the idea can be extended to such combinations as $29 + 6 = 35$, $49 + 8 = 57$, and so on, and then to $25 + 9 = 34$, $67 + 9 = 76$, and so on. Still later, perhaps, when he has begun to do column addition he learns that

1

The Growth and Development of Mathematical Ideas in Children

Or How and Why To Use This Book

PHILLIP S. JONES

WE WHO are planning and writing this book have on occasion disagreed as to the relative importance of some of the goals of mathematics education, and as to the best methods for achieving them, but we are agreed that the following are axioms:

AXIOM 1. *The best learning is that in which the learned facts, concepts, and processes are meaningful to and understood by the learner.*

AXIOM 2. *Understanding and meaningfulness are rarely if ever "all or none" insights in either the sense of being achieved instantaneously or in the sense of embracing the whole of a concept and its implications at any one time.*

The sudden perception or flash of insight which is one of the joys of mathematics learning and teaching comes only to those who have, with thought, struggled to extend or apply concepts which have been partially understood earlier. Further, meanings and understandings themselves change continuously as they are extended, broadened, and applied in different situations. For example, the early elementary pupil who suddenly perceives that 9 plus any other single digit number is the teen number whose units digit is one less than the single digit number probably first sees this after having added a number of specific combinations such as $9 + 3 = 12$, $9 + 4 = 13$, $9 + 6 = 15$, and so on. Even if he had been told this rule, it is doubtful that it would have been really meaningful to him until after he had worked out several examples. Later the idea can be extended to such combinations as $29 + 6 = 35$, $49 + 8 = 57$, and so on, and then to $25 + 9 = 34$, $67 + 9 = 76$, and so on. Still later, perhaps, when he has begun to do column addition he learns that

addends may be regrouped without changing the sum. Thus the sum

$$\begin{array}{r} 3 \\ 4 \\ 6 \\ 8 \\ \underline{2} \end{array}$$

may be thought through mentally either as $3 + 4 = 7$, $7 + 6 = 13$, $13 + 8 = 21$, $21 + 2 = 23$, or as $3 + (4 + 6) + (8 + 2) = 3 + (10 + 10) = 3 + 20 = 23$. This latter may seem to be different in principle from the rules for adding nines, but both are merely special cases of the *associative law* which, in algebra, is written $a + (b + c) = (a + b) + c$. However, even before studying algebra, and without any use of the term *associative law*, the student can work out the reason for his rule for adding nines by thinking: $9 + 3 = 9 + (1 + 2) = (9 + 1) + 2 = 10 + 2 = 12$ or $29 + 6 = 29 + (1 + 5) = (29 + 1) + 5 = 30 + 5 = 35$. When all of this has happened, his understanding of an important idea has grown and developed from being used and understood at a very simple nonverbal level to the recognition of a broadly applicable principle with a name (associative law or axiom).

The actual sequence of extensions and insights will vary from pupil to pupil and teacher to teacher. Our point is that the ideas should be understood at each stage, but that the nature and generality of the pupil's conception will also grow and develop and should not in general be expected to be complete at any stage. In fact, the associative law, the use and understanding of which has its roots in the early elementary years, is an increasingly important principle in many mathematical theories, such as vectors, matrices, and linear algebras, still to be met by the student after he has completed secondary school.

The two axioms with which we began imply two theorems and a couple of corollaries. They are:

THEOREM 1. *Teachers must plan so that pupils continually have recurring but varied contacts with the fundamental ideas and processes of mathematics.*

These contacts should often be in different contexts. They should be both at higher levels of abstraction and generalization and also in the form of concrete applications or realizations of old generalizations in new situations. In both new generalizations and new specific cases or applications it is important that the student be deliberately led to see the continuing theme, the general principle, which was met earlier and now is being extended or applied. For example the child who has learned that $3 \times 10 = 30$, $3 \times 2 = 6$ may then be lead to think for himself

that $3 \times 12 = 3(10 + 2) = (3 \times 10) + (3 \times 2) = 30 + 6 = 36$ before he learns the mechanical process or algorism represented by

$$\begin{array}{r} 12 \\ \underline{3} \\ 36 \end{array}$$

The fact that $3 \times 12 = 3 \times (10 + 2)$ could have been developed out of such still earlier concepts as that multiplication by an integer, 3, is equivalent to repeated addition, and that 12 means $10 + 2$. Thus, if 12 were represented by two piles of apples, one with 10 apples and the other with 2, three times this quantity could be represented by three piles of 10 each and three of two each, or a total collection of 36. Later the student meets this same principle when he learns that three packages containing 1 lb. 4 oz. each will together contain 3 lb. 12 oz. Again, geometrically, the area of the two rectangles in Figure 1 is $3 \times (5 + 4) = (3 \times 5) + (3 \times 4) = 15 + 12 = 27 = 3 \times 9$ sq. ft., and in studying angular measure in either geometry or trigonometry $3 \times 15°13' = 45°39'$. This same principle which he may, in algebra, learn to write as $a(b + c) = ab + ac$ and call the *distributive law* will help him to understand why a negative number times a negative number is defined to be a positive number. It also, of course, is the concept underlying factoring in algebra and the processes of multiplication and division in arithmetic and algebra.

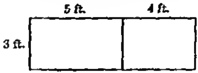


FIG. 1

As each of these topics is studied the teacher should point out the general principle being used, its occurrence in earlier situations, and exactly how it is being used or extended now.

Displaying such examples of the continuing growth and extension of basic mathematical ideas is the function of the rest of this book, not of this chapter. To point out how and why to read the rest of this book we need one more theorem.

THEOREM 2. *Teachers in all grades should view their tasks in the light of the idea that the understanding of mathematics is a continuum, that understandings grow within children throughout their school career.*

This theorem implies immediately the corollaries that (1) Teachers should find what ideas have been presented earlier and deliberately use

them as much as possible as the basis for the teaching of new ideas. (2) Teachers should look to the future and teach some concepts and understandings even if complete *mastery* cannot be expected.

Teachers should do this in order to build *readiness* for new topics. We have all at sometime had a flash of insight or perception of relationships come to us as we went over a topic a second time in a later course or in a new context. This growth in understanding could not have come the second time over if there had not been a first time! We must all be content, even be anxious, to be the unsung heroes of some later teacher's triumph. Let's hope we also take care to reap the fruit of the seeds sown by our predecessors!

Careful consideration for the *vertical* organization and continuity of our mathematics program not only in its content but even in its presentation, especially in its emphasis on the understanding of basic ideas, will not only tend to create readiness for new learnings, but will tend to eliminate possible sources of later interference. The elementary teacher, who is aware that in the eighth or ninth grade his pupils will learn of the negative numbers which make subtraction always possible, will be a little cautious about insisting that one can never subtract a larger from a smaller number. Early stress on subtraction and division as, by definition, the inverses of addition and multiplication not only introduces preliminary ideas of proof and of the structure of mathematical systems, but also will help teachers to lead students to understand why one can not divide by zero as well as why one "*changes the sign* and *adds*" and "*inverts and multiplies*" in certain well-known situations.

Theorems 1 and 2 and their corollaries indicate the readers who we hope will find value in this book and the uses to which we hope it will be put.

We hope this book will be valuable to all teachers of mathematics, K through 12. Since it is clear that these teachers do not all have the same background nor the same problems, we have occasionally dispaired of writing in one book materials which would be understandable by and useful to all teachers. However, we are so convinced of the importance of continuity in the mathematics program as a whole that we decided to attempt the task.

We do *not* believe, for example, that *proof* plays the same important role in the first grade as does *number*. However, in the early elementary grades children do *prove* that $7 - 3 = 4$ by showing that $4 + 3 = 7$, and a third grader can reason that $8 + 5 = 13$ *because* 5 is $2 + 3$, $8 + 2$ is 10, and hence $8 + 5 = 10 + 3 = 13$. Thus, his introduction to the nature of and need for proof and for reasons can actually be begun in his early study of number.

So readers, we urge all of you to follow through the book with us. You who are *elementary school teachers* will find in each chapter some discussions which apply to your grades. These are, in general, presented early in each chapter. We hope you will read a little further than this in each chapter in order to see how the structure of mathematics, not merely of computation, may be based upon the foundations which you lay. However, when the discussions of number ideas seem clearly to have moved far beyond your instructional level, turn to "Relation and Function," "Proof," and so on, to see what these may hold for you.

You who are *junior and senior high school teachers* can scan our essays for connections with earlier work which may be emphasized in your teaching, as well as for how you, too, may help basic mathematical understandings grow in the minds of your students. *Read both what may come before and what may come after your present level of instruction.*

And finally, may we say a few words about what is not here? Just as chapters on "Proof" and "Probability" are included even though they may be central ideas which should be and are less stressed at the early elementary level than later, so, too, there certainly are important ideas and concepts which belong in the mathematics program for which we didn't find any room in this book. We selected what to us seemed the most important continuing concepts but we do not pretend that the book covers everything that should be taught.

However, when you seem to miss *algebra* or *geometry* do not assume that we have relegated them to an unimportant role in the school of the future. Look through "Number and Operation," "Relations and Function," even "Proof," "Measurement and Approximation," and "Language and Symbolism in Mathematics." You will find algebra in all these chapters and in others too. Similarly geometry may be found in "Proof," "Relations and Function," and "Measurement and Approximation." It is our belief that continuity, growth, and development of understanding may be planned for and achieved in many curricular contexts both traditional and nontraditional. You need not have an integrated curriculum to do this, though some will think it easier to plan in such a setting. However, the central themes which we wish to stress are not revealed or emphasized by course titles such as *algebra*, *geometry*, and *trigonometry*. Hence we have not used these courses for the framework of our book. We believe that the major concepts of these courses fall under one or more of the central concepts which we have discussed. However, space has not permitted us to expand these discussions to display everything which may be classified with each concept or theme. We hope you will read in each chapter, and that you will seek

and see in it the topics you are now teaching (plus perhaps a few more) from a new viewpoint.

Further, we recognize that at least two of our chapters, "Probability" and "Statistics," represent topics which certainly are not now accorded a central or unifying role in our elementary and secondary school curricula. We included them because we feel certain that recent rapid developments in both their theories and applications justify us in claiming for them a more central position in the curriculum of the future. We did not set out to prophesy or to advocate changes, but felt it to be proper, even to be a duty, to recognize changes which seemed desirable to all of us. We have tried to make it clear as we went wherever our prophecies have outrun our experience.

Don't skip merely because some new terms or symbols are being used—the underlying ideas may be old and familiar ones which new terms and symbols may simplify. Similarly, don't skip because the material being discussed is old stuff—follow along in search of some new aspects of teaching.

Scan the "Preface" for another statement of our objectives, read at least part of *each* chapter, and finally turn to the last chapter for suggestions as to how to extend further your study of these ideas and their use in both teaching and curriculum planning. We do *not* believe there is just one correct sequence or grade placement of topics. We are concerned that every teacher be aware of his position in and contribution to a planned spiral development of ideas in his school. Thus the "Flow Chart" in Chapter 11 is *not* a grade placement chart nor do we regard it as defining a correct or rigid sequence. It is intended as an example of a helpful analysis which might profitably be thought through for itself by the staff of every school system.

We hope you will also observe that there are some themes or recurring concepts running through this book. For example, even the most elementary mathematics, counting, is an abstraction. Abstractions and their related generalizations are what make mathematics useful—the same principles apply in many situations. In teaching, one should plan to complete the cycle, moving from concrete to abstract to concrete realizations, applications, or models of a theory. Move from special to general, and back to further specializations. New ideas are often *discovered* inductively and/or by seeking new extensions which will still preserve the essential properties of the old system from which they stem. But let's move on and see what *you* find as themes for our book!

2

Number and Operation

E. GLENADINE GIBB, PHILLIP S. JONES,
AND CHARLOTTE W. JUNGE

"Mathematics is the queen of the sciences, and arithmetic the queen of mathematics."—C. F. GAUSS

"Reeling and Writhing, of course, to begin with, ... and then the different branches of Arithmetic—Ambition, Distraction, Uglification, and Derision."—LEWIS CARROLL (C. L. DODGSON) in *Alice's Adventures in Wonderland*

A DAY without number! ... how often children are asked to speculate on such an event in an attempt to envisage the importance of mathematics. What would happen if there were no numbers? What changes would come to some of the things we take for granted in everyday living? No prices would appear on the food we wish to buy at the corner grocery. We would need to determine other ways of recording time for all numerals on the calendar, and the clock would disappear. We would find ourselves perplexed in trying to figure out a way to report business transactions. Our latest scientific developments would become extremely difficult if not impossible to understand or use. Yes, probably the oldest unifying theme in mathematics is that of number.

Number is both a familiar idea and a fundamental idea to all of us. In fact, it is so much a part of our everyday lives that we may easily lose sight of the nature of its development. In this chapter it is our purpose (1) to display the fundamental ideas of number and operation and (2) develop with our readers an appreciation for these fundamental ideas of number and operation as they exist in the structure of mathematics, not only from a viewpoint of logical structure but also from a viewpoint as to how these ideas develop in the minds of students. More specifically, we shall examine the nature of number, how the natural numbers are used to invent new numbers—the fractions, the positive and negative numbers, the irrational numbers, and the complex num-

bers. As we proceed, we shall follow the order of the major steps in which new numbers are introduced today in our schools. We also shall consider the fundamental operations of arithmetic—addition, subtraction, multiplication, and division—with each kind of number.

The development of mathematics may be regarded as the development of two central themes, namely, *number* and *form*.[1*] By the latter we refer not only to geometric form, but also to the structure of mathematical relations and theories as developed in sequences of postulates and theorems. These sequences and their accompanying proofs are found not only in geometry, but also are logically necessary in algebra and all other divisions of mathematics, all of which in turn have their foundations in elementary school arithmetic.

In the seventeenth century these two themes, number and form, were integrated to form mathematical analysis, the analytic geometry, and the calculus. More recently, in the latter part of the nineteenth century, we also have the intermingling of these two themes, number and logical structure or form, in an attempt by Peano[2] and others to axiomatize not only natural numbers but all numbers.

As man has encountered new problems, he has created a need for new numbers. Consequently, over the ages he has successively invented new numbers to satisfy new needs. At each step, old numbers have been the building blocks for new numbers. His guide in using these building blocks has been that of preserving such basic properties of old numbers as order and equivalence and the commutative, associative, and distributive laws of operation with these numbers. Although some modern algebras and their arithmetics fail to have all of these properties, such properties will continue to be fundamental in elementary and secondary school mathematics. This is true not only because numbers with these properties were the first to develop historically, but also because these mathematical properties correspond to the way in which most of the objects of our immediate physical world behave. For example, the number of apples in a pile is the same whether the pile was formed by putting down first three apples and then two apples, or first two apples and then three apples.

We believe that three major considerations should be reviewed by a teacher in planning for maximum meaning and understanding. They are (1) *logical considerations* which are based on the structure and logic of mathematics itself, (2) the *considerations of history—the chronological sequences* of events which sometimes furnish the only real reason for the existence of particular terms or units being taught, such as root, inch,

* A superscript in the text refers the reader to bibliographical references at the end of the chapter.

degree, and (3) the *pedagogical sequence*—a sequence based on sound psychological considerations whereby the teacher weighs such things as the necessity of proceeding from the use of objects as representations of ideas to abstractions, to generalizations, and then returning to apply the generalizations to specific situations. Consideration should also be given to levels of learning in moving from less mature to more mature ideas.

No longer do we confront a child with a new idea by using the adult level of a written algorism and expecting him to memorize it. Rather, we recognize that the development of mathematical concepts is a growth process going from levels of working with manipulative materials through various levels of working with abstract symbols. For example, a child in the lower elementary grades may be able to determine by using manipulative materials how many candies he can buy for fifteen cents if he knows that three candies cost five cents. By putting three objects with five objects, another three objects and another five objects, and another three objects with another five objects he finds that he can get nine candies for fifteen cents. Later, in the upper elementary school, he may express this as

candies	3	?
cents	5	15

or, $\frac{3}{5} = \frac{?}{15}$.

From his understanding of classes of equivalent fractions, he knows that $\frac{3}{5}$ is another name for $\frac{9}{15}$, and therefore, the number of candies must be nine. Still later in the secondary school, he may express this situation as

$$\frac{3}{5} = \frac{n}{15}$$

but he solves the problem by determining the number with which to replace n in the following manner:

$$5n = (3)(15)$$
$$5n = 45$$
$$n = 9.$$

Sometimes, the going from one level to another is almost instantaneous. Again, it may take several years as noted in the preceding illustration.

WE LEARN NUMBER NAMES

Logical, Chronological, and Pedagogical Approaches. These approaches to the development of numbers all tend to indicate a similar type of teaching sequence and method. Seemingly, the integration

of these considerations into a sound classroom procedure is easier for the development of the concept of number than it is in some later areas where a parallelism is not so obvious or may not even exist. To see this parallelism, let us consider simultaneously the number experiences of prehistoric man and the preschool child.

It is believed from recent studies of anthropologists and excavations of archeologists that prehistoric man would keep a record of the number of objects, such as skins or beads, by making tally marks on a stick of wood, or in the case of the earliest known mathematical writing, on a leg bone of a prehistoric wolf. This prehistoric bone, unearthed in Moravia in 1938,[3] bears on it scratches in an orderly array—a group of five, a second group of five, and so on. After five such groups of five scratches, a longer notch appeared, followed by a continuation of subgroups of five until a total of fifty-five scratches had been recorded.

What were the purposes of these scratches? No one knows exactly nor are we certain that prehistoric man was keeping a record of his possessions by tallying. The chances, however, that these would have been arranged in this fashion by accident are so slight and the connection with tallying with notched sticks and knotted ropes so clear that there is little doubt that some prehistoric man was working with number ideas. Primitive people have been observed to make a one-to-one correspondence between sheep, skins, wives, children, and collections of pebbles or notches on such recording surfaces as sticks or bones of animals.

Studies[4,5,6,7] have shown that number has been a part of the experiences of many preschool children. These have been for the most part informal experiences. A child holds up three fingers to report his age. He sees that there is a plate on the table for each member of the family. He speaks of the four candle-holders and the four candles to be placed in the holders. He learns to determine each move in a game of parchesi by counting the spots on a die. He keeps his score in a dart game by making tally marks on the playroom blackboard.

From brief accounts of the beginnings of number in our culture, we note the similarity of prehistoric man and the preschool child in their establishing a one-to-one correspondence between two sets of objects. Later, a set of marks or symbols replaced one of the groups of things. This progress from things to symbols and the use of correspondences suggests an approach desirable in the development of the idea of number within classroom situations.

More specifically, one of children's early experiences with number comes with identifying the concept of *two* common to his two hands,

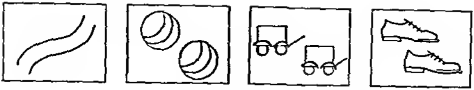
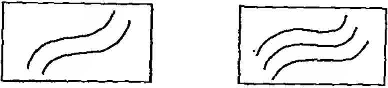


FIG. 1



A B

FIG. 2

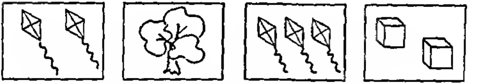


FIG. 3

his two eyes, his two feet, his two ears. He recognizes a *one-to-one correspondence* between the elements in the sets shown in Figure 1. *This common property of these groups or sets* is given the name, '2'.* He learns that there is not a one-to-one correspondence between elements in such sets, as shown in Figure 2. The cardinal number of the set in Figure 2 A can be given the name '2'. The cardinal number of the set in Figure 2 B cannot be given the name '2'. It is given another name, '3'. Having arrived at some understanding of what sets have the cardinal number 2, the child is able to select the collections whose number may be appropriately labeled by the symbols 'two' or '2' from such groups as illustrated in Figure 3.

The Concept of Set and Cardinal Number. *In referring to the pictures of groups of objects in the preceding section, we have called them 'sets'.* One of the most fruitful ideas for the mathematician in recent years has been the concept of set. Regarded as an undefined term in mathematics, the word 'set' is synonymous with such words as 'collection', 'aggregate', 'menge' (German), and 'ensemble' (French). In effect,

* Single quotation marks will be used to denote the numeral or the name of the number.

we can say that a set is a collection of things. The objects making up the set are said to be members or elements of the set. Some scheme must be given so that we may know whether or not a particular element is a member of that set. For example, the set may be all the elementary teachers in Lincoln School; it may be the set of all presidents of the Mathematics Club since its founding; it may be the set of all students in Mr. Smith's algebra class at West High where Nancy White is not a member of the set since she is in Miss Dunn's class; it may be the set of all even integers whose members we can determine by merely testing whether or not the integer is divisible by 2; it may be the set of natural numbers from 1 through 10. This set may be represented by $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. We can easily tell that 3 and 4 are members of this set while 12, $3/5$, and $5/3$ are not members of the set. It may be the set of all medians of a given triangle; or, it may be the set (an infinite set, incidentally) of all pairs of numbers, x and y, such that they would satisfy the condition that $x + y = 8$.

Two sets are said to be *equivalent* if for every element of one set there is one and only one element of the other set; that is, if for given Sets A and B, there exists a one-to-one correspondence between the individual elements of Set A and of Set B (Fig. 4). Equivalent sets are said to have the same *cardinal number*; e.g., the cardinal number of Set A is 3 and the cardinal number of Set B is 3. In dealing with equivalent sets and their cardinal numbers we disregard the specific nature of the objects being matched whether it be a set of pebbles matched with a set of sheep, a set of plates with a set of people, or a set of symbols with the set of fingers of a hand. We are concerned only with the common property of the sets—known as their cardinality—their fourness or their fiveness. If the sets are not equivalent, they lack common cardinality.

Set A Set B

FIG. 4

The equivalence of two sets may be demonstrated in several different ways. For example, we can show that a set of balls and a set of blocks are equivalent by matching as shown in *a* or *b* of Figure 5. Either of these schemes shows that for every element of one set, there is one and only one corresponding element of the other and consequently shows that the two sets are *equivalent*.

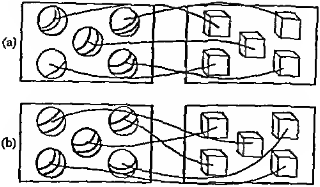


FIG. 5

Let us consider the sets in Figure 6, labeled A, B, and C. Although Set A is a set of toys, Set B is a set of fingers, and Set C is a set of letters, each of the elements of one set may be placed in one-to-one correspondence with an element of each of the other sets; that is, each element in Set A may be matched with an element in B and with an element in C, as illustrated in Figure 7. Similarly, each element in Set B

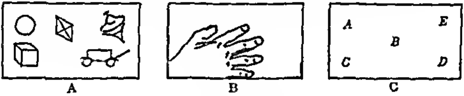


FIG. 6

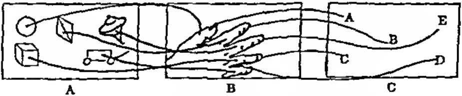


FIG. 7

may be matched with an element in Set C and with an element in Set A. Since the elements of these sets may be put into one-to-one correspondence, they are said to have the same cardinal number. This comparison of sets of elements with respect to cardinality is possible without the use of number names or counting. For example, we may see that each person in a given set of people matches with a chair in a given set of chairs and yet be unable to say how many chairs or how many people there are in either set. If there are some empty chairs after everyone is

seated, we know that we have more chairs than people but we may not be able to say how many of each.

It is convenient, however, to have names for each of the cardinal numbers. These names may be anything we wish. For instance, we call the cardinal number of Sets A, B, and C, 'five' because historically this word was invented, used, and understood long before the importance and usefulness of this approach to cardinal numbers by way of equivalent sets was recognized. Yet, this number could have been given any arbitrary name, such as, *'lu'*, *'bee'*, *'zoke'*, *'zume'*, *'funf'*, *'cinq'*.

THE COUNTING NUMBERS—1, 2, 3, ···

The Natural Numbers. Throughout the elementary and secondary schools, we are concerned primarily with the counting numbers. They are man's way of indicating how many objects are contained in a given collection. These counting numbers, the simplest of numbers, 1, 2, 3, ··· 11, 12, 13, ··· 101, 102, 103, ··· 1001, 1002, 1003, ··· are the building blocks for the other numbers—fractions, positive and negative integers, irrational numbers, and complex numbers. Other names commonly used for these counting numbers are 'integers', 'whole numbers', 'cardinal numbers', and 'natural numbers'. In mathematical usage, as opposed to common usage, there is a distinction in these names. The integers include both positive and negative numbers and zero. Just as we termed *equivalent* all finite sets that can be put into one-to-one correspondence with each other and have associated a number name with each set of all such sets, so infinite sets that can be placed in one-to-one correspondence are said to be equivalent and to have the same cardinal number. For example, we can place in one-to-one correspondence the set of even numbers and the set of positive integers as

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \cdots n \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ 2 & 4 & 6 & 8 & 10 & 12 & \cdots 2n \cdots. \end{array}$$

All sets which can be put into one-to-one correspondence with the set of positive integers are equivalent infinite sets and are given the cardinal number *aleph-null*, $\aleph_0$. Aleph-null is the first or least 'transfinite' cardinal number.

The term, 'cardinal number', is applied to both finite and transfinite numbers. The technical term that means only counting numbers is 'natural numbers'. We shall use this term when referring to the counting numbers, the numbers that are commonly called whole numbers in elementary school mathematics.

Early Number Experiences. Children's early number experiences may follow either an ordinal or a cardinal approach. Using an ordinal approach, a child learns the number names in sequence—one, two, three, four, $\cdots$. He learns that the name three comes after two, that eight comes before nine and after seven. He may then mechanically match a sequence of number names with a sequence of objects. When he is counting rationally, that is, with understanding and insight into the counting process, he realizes that the last name used is the name of the cardinal number of the entire set. He knows that the arrangement of the objects does not make a difference as long as each object is matched with a number name in the ordered sequence of cardinal numbers.

Using the cardinal approach, one of a child's early experiences in a study of number is to recognize and learn the natural numbers as names for standard sets or groups. We shall follow from this approach. A child identifies the following collections as among those which may be given the name 'five', and learns to record the common property with the numeral, '5'. Any one of the Sets, A, B, or C in Figure 7 or $\{/////\}$ or even $\{6, 2, 8, 4, 7\}$ may be used as a standard set for 5 and there need be no previous knowledge of such symbols as '1', '2', '3', '4', or of an order for these symbols in order to obtain the cardinal number, 5. However, it also would be possible to use the previously memorized, ordered set of symbols $\{1, 2, 3, 4, 5\}$ as the *representative set* for defining the cardinal number, 5.

Ordering Standard Sets. To continue to define so-called standard sets independently of each other would result in a long list of names to be associated with a corresponding number of standard sets. To determine the cardinal number of a set it would be necessary to try to establish a one-to-one correspondence between the given set and the elements of many standard sets until a standard set was found such that each element in the standard set was in one-to-one correspondence with each element in the given set. This soon becomes a most inconvenient way of determining the cardinal number of a given set.

This can be done more easily by considering an additional property of the natural numbers, a property that is related to the very feature which has made them seem *natural*. We shall let the sets in Figure 8 be

FIG. 8

representative sets for 1, 2, and 3 respectively. We could soon have an assorted array of standard sets, as illustrated in Figure 9.

If we compare any two of these sets by matching elements, either they are equivalent or some elements of one set are *left over*.

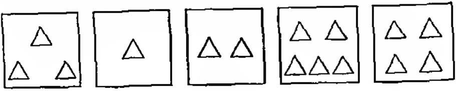


FIG. 9

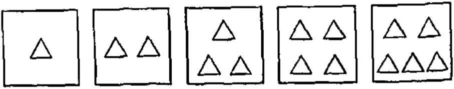


FIG. 10

Using this idea of nonequivalence, we can arrange an ordered sequence of models as in Figure 10 where the successor of a given set is a set equivalent to the given set with one more element in it. We then can assign names and symbols to the cardinal numbers of these sets. These cardinal numbers are ordered by the ordering of the representative sets. Thus, 1 is less than 2 is less than 3 because 2 is $1 + 1$, 3 is $2 + 1$, ... n is $(n - 1) + 1$. Once this ordered system is created and the order memorized, *counting a collection* means assigning to every member of the set a term in the ordered sequence of number symbols. If any set of discrete elements is ordered in some way, we also can use the ordered sequence of numerals to describe the position or place of a particular element in its set. For example, John sits in the fourth row from the windows in the third seat from the front of the room. Or, in Figure 11 the third ring from the left may be identified as illustrated.

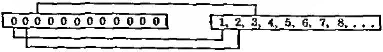


FIG. 11

The ability to set up correspondences in this specialized way comes into existence when the first few number words have been committed to memory in their ordered succession and a scheme (including a system of numeration) has been devised to pass from any natural number to its successor. Each new natural number after 1 may be derived from

its predecessor by adding 1. Without our ability to arrange things in ordered succession, little scientific progress could have been made. The last element in the set of numerals {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...} needed to match with the set of objects names the number of the set of objects or elements. By this matching process, we are able to answer two questions—"How many?" and "Which one?" Children may learn these number names with an understanding of the sets they label before they can read the name or recognize the numeral. Furthermore children need not know the words 'cardinal' and 'ordinal' to understand and use these ideas.

THE SYSTEM OF NUMERATION

Early Systems. From the earliest times, most civilizations have had some system of numeration. At first these systems were unrecorded for men developed mathematical understandings and used mathematical ideas long before they had any methods of recording these ideas. Manipulative materials (pebbles, sticks) and mechanical counting and calculating devices (abacus, counting board) were used to keep records of possessions and transactions.

These early numeration systems, while varying one from another in form and complexity, also revealed a high degree of similarity in terms of basic principles. Each involved the use of a base, a way of grouping. Some systems were based on five, the Egyptians and Romans used a base of ten, the Mayas of Yucatan used the base of twenty, the Babylonian system was based on sixty, and some primitive tribes near Australia had a base of two. For the Egyptians, Greeks, and Romans, the number represented by a collection of number symbols was essentially independent of the position or sequence in which the symbols were written. Remnants of these various systems are seen in our number system today, such as the use of *score* for 20 and the division of hours and degrees into 60 minutes, and the Roman numerals which are still used for such purposes as inscriptions on public buildings, numerical tablets, cornerstones, ornamental clock faces, and for designating chapters of books.

From the early *unwritten* numeration system, the present decimal numeration system has evolved. While it is called 'Hindu-Arabic', it is in reality a composite of ideas from many civilizations, including the Babylonians and the Greeks. It cannot rightly be claimed in its entirety by any particular group. Using the ideas of symbol, base, and positional notation, we often gain further insight into our decimal system by using other bases instead of ten. For example, using the symbols to which we

representative sets for 1, 2, and 3 respectively. We could soon have an assorted array of standard sets, as illustrated in Figure 9.

If we compare any two of these sets by matching elements, either they are equivalent or some elements of one set are *left over*.

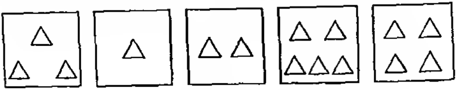


Fig. 9

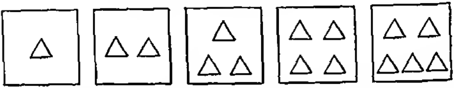


Fig. 10

Using this idea of nonequivalence, we can arrange an ordered sequence of models as in Figure 10 where the successor of a given set is a set equivalent to the given set with one more element in it. We then can assign names and symbols to the cardinal numbers of these sets. These cardinal numbers are ordered by the ordering of the representative sets. Thus, 1 is less than 2 is less than 3 because 2 is $1 + 1$, 3 is $2 + 1$, . . . n is $(n - 1) + 1$. Once this ordered system is created and the order memorized, *counting a collection* means assigning to every member of the set a term in the ordered sequence of number symbols. If any set of discrete elements is ordered in some way, we also can use the ordered sequence of numerals to describe the position or place of a particular element in its set. For example, John sits in the fourth row from the windows in the third seat from the front of the room. Or, in Figure 11 the third ring from the left may be identified as illustrated.

Fig. 11

The ability to set up correspondences in this specialized way comes into existence when the first few number words have been committed to memory in their ordered succession and a scheme (including a system of numeration) has been devised to pass from any natural number to its successor. Each new natural number after 1 may be derived from

its predecessor by adding 1. Without our ability to arrange things in ordered succession, little scientific progress could have been made. The last element in the set of numerals {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...} needed to match with the set of objects names the number of the set of objects or elements. By this matching process, we are able to answer two questions—"How many?" and "Which one?" Children may learn these number names with an understanding of the sets they label before they can read the name or recognize the numeral. Furthermore children need not know the words 'cardinal' and 'ordinal' to understand and use these ideas.

THE SYSTEM OF NUMERATION

Early Systems. From the earliest times, most civilizations have had some system of numeration. At first these systems were unrecorded for men developed mathematical understandings and used mathematical ideas long before they had any methods of recording these ideas. Manipulative materials (pebbles, sticks) and mechanical counting and calculating devices (abacus, counting board) were used to keep records of possessions and transactions.

These early numeration systems, while varying one from another in form and complexity, also revealed a high degree of similarity in terms of basic principles. Each involved the use of a base, a way of grouping. Some systems were based on five, the Egyptians and Romans used a base of ten, the Mayas of Yucatan used the base of twenty, the Babylonian system was based on sixty, and some primitive tribes near Australia had a base of two. For the Egyptians, Greeks, and Romans, the number represented by a collection of number symbols was essentially independent of the position or sequence in which the symbols were written. Remnants of these various systems are seen in our number system today, such as the use of *score* for 20 and the division of hours and degrees into 60 minutes, and the Roman numerals which are still used for such purposes as inscriptions on public buildings, numerical tablets, cornerstones, ornamental clock faces, and for designating chapters of books.

From the early *unwritten* numeration system, the present decimal numeration system has evolved. While it is called 'Hindu-Arabic', it is in reality a composite of ideas from many civilizations, including the Babylonians and the Greeks. It cannot rightly be claimed in its entirety by any particular group. Using the ideas of symbol, base, and positional notation, we often gain further insight into our decimal system by using other bases instead of ten. For example, using the symbols to which we

TABLE 1

Set of elements .	●	●	●	●	●	●	●	●	●	●	●	●
Set of numerals (base ten) .	1	2	3	4	5	6	7	8	9	10	11	12
Set of numerals base seven)...	1	2	3	4	5	6	10	11	12	13	14	15

are accustomed and counting on the base of seven, we determine the cardinal number of the set of elements by pairing them with number symbols as illustrated in Table 1.

In other words the same number is represented by 12 in the decimal system and by 15 in a numeration system based on seven.

Table 2 illustrates the use of the base of two. If we use the duodecimal system (base of twelve) we must add two new symbols for the '10' and '11' of the decimal system. Letting these symbols be '*t*' and '*e*', we count as illustrated in Table 3.

TABLE 2

Set of elements..	●	●	●	●	●	●	●	●
Set of numerals (base ten)	1	2	3	4	5	6	7	8
Set of numerals (base two)	1	10	11	100	101	110	111	1000

TABLE 3

Set of elements	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Set of numberals (base ten). .	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Set of numerals (base twelve)...	1	2	3	4	5	6	7	8	9	*t*	*e*	10	11	12	13

For a natural number written in the duodecimal system, '13' represents 1 twelve and three which corresponds to 15 (1 ten and 5 ones) in the decimal system, to 33 on the base of four (3 fours and 3 ones) or 30 on the base of five (3 fives and 0 ones).

Why then do we have the base of ten? It is strictly an inheritance from ancient times. Had we had twelve fingers instead of ten, we have reason to believe that the duodecimal system would today be in world-wide use among civilized people.

Our System of Numerals. The Hindu-Arabic system of notation, like most great achievements, is relatively simple. It is an additive system, utilizing (1) ten symbols ('0', '1', '2', '3', '4', '5', '6', '7', '8', '9'), and (2) positional notation. Each numeral is a name of a cardinal number in its own right and represents a standard set. Furthermore, each numeral has another role. By its position in the name of an integer, it indicates the size of the subsets which it enumerates: '3' is the numeral for a set of three elements; '30' is the numeral for a set of 3 collections of 10 each or of thirty elements.

Let us consider the numeral, 6666. The digit, '6', represents 6 of something. The place in which the '6' is written indicates whether the *something* is ones, tens, hundreds, and so on. An example is illustrated in Scheme 1.

SCHEME 1

6 6 6 6

- 6 ones or 6×1 or 6×10^0 ($10^0 = 1$ by definition)
- 6 tens or 6×10 or 6×10^1
- 6 hundreds or $6 \times 10 \times 10$ or 6×10^2
- 6 thousands or $6 \times 10 \times 10 \times 10$ or 6×10^3

As one progresses from right to left across the numeral, '6666', the number represented by each '6' is ten times the value of the one preceding it, or, if one reads from left to right, each succeeding digit represents one-tenth the value of the one preceding it. Thus, the cardinal number represented by any digit is dependent upon the digit itself *and* the place it occupies.

These same principles, operative in natural numbers can be extended to decimal fractions. In the numeral '1.1', the value of '1' in the 'tenths' place is one-tenth the value of that in the 'ones' place. Decimal fractions also compound in powers of ten as do the natural numbers. Let us now consider the numeral, '6666.666' as illustrated in Scheme 2.

SCHEME 2

6 6 6 6 . 6 6 6

- 6 thousandths or $6 \times \frac{1}{1000}$ or $6 \times \frac{1}{10^3}$ or 6×10^{-3}
- 6 hundredths or $6 \times \frac{1}{100}$ or $6 \times \frac{1}{10^2}$ or 6×10^{-2}
- 6 tenths or $6 \times \frac{1}{10}$ or $6 \times \frac{1}{10^1}$ or 6×10^{-1}
- 6 ones or 6×1 or 6×10^0
- 6 tens or 6×10 or 6×10^1
- 6 hundreds or 6×100 or 6×10^2
- 6 thousands or 6×1000 or 6×10^3

OPERATIONS WITH NATURAL NUMBERS

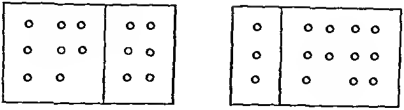
Addition. Early in his study of number a child sees that the collection of 6 is the same as a combination of collections of 3 and 3, of 4 and 2, of 5 and 1, and so on. Later, he becomes familiar with the sign ('=') between numerals and learns that the collection designated by the symbol on one side of the sign is equivalent to the collections designated by the symbols on the other side. For example, a set of balls whose number is 4 and another set of balls whose number is 2 may be combined to form a set of balls whose number is 6 ($4 + 2 = 6$). This bringing two sets or collections together is known as addition. It results in the same pattern that the child recognized in his study of the group of 6 where he learned that '$4 + 2$' is another name for 6.

At advanced levels in the understanding of this same idea, in the secondary school and junior college, we may say that if we have two sets, A and B, we can consider a third set which contains every element which is in either A or B. This new set is called the *union* of A and B ($A \cup B$). If A and B are finite, disjoint sets, that is, nonoverlapping such that no element may be in both sets, and have cardinal numbers 'a' and 'b' respectively, then we can denote the cardinal number of the set $A \cup B$ by '$a + b$'. The operation that determines this number, $a + b$, is addition. Consequently, we may come to think of '$6 + 2$' as a number, as merely another name for '8', and not as an operation of addition as thought of earlier in the elementary school.

It is important for the development of number sense or understanding that children practice thinking of the single numbers as representing many different pairs. For example, a child may think of 7 as $3 + 4$, $5 + 2$, $6 + 1$. He later uses this idea in adding $6 + 7$ by thinking $6 + 7$ is the same as $6 + (4 + 3)$ and then as $(6 + 4) + 3$, as $10 + 3$, and finally as 13.

This process of regrouping $6 + (4 + 3)$ as $(6 + 4) + 3$ is seen to be an instance of the *associative law* and is perceived eventually as a basic common property of all elementary number systems in arithmetic, algebra and even vector algebra. More generally, if we should have three nonoverlapping or disjoint finite sets, A, B, and C, with a, b, and c, the cardinal numbers of the sets, then the cardinal number of the union of these three sets will be the number $(a + b) + c = a + (b + c)$, depending on the order in which we conceived of the sets to be combined.

Diagramically, letting $a = 3$, $b = 5$, and $c = 6$, the union of Set A, Set B, and Set C may appear as illustrated in Figure 12A or B. Since the final new set and the elements in it will be the same in either case, we have $a + (b + c) = (a + b) + c$ which is the associative law of


$(a + b) + c$
A

$a + (b + c)$
B

FIG. 12

addition. It simply means that in adding 3, 5, and 6, one can add 3 and 5 and to that sum add 6, or to the sum of 5 and 6, add 3.

A child also learns that putting 2 balls with 4 balls gives him the same results as putting 4 balls with 2 balls ($4 + 2 = 6$ and $2 + 4 = 6$). The recognition of this property is very helpful to children in learning number facts, for if they know that $3 + 2 = 5$, then they also know that $2 + 3 = 5$. Given such exercises as

$$\begin{array}{r}32\\+\underline{21}\end{array}\qquad\begin{array}{r}23\\+\underline{21}\end{array}\qquad\begin{array}{r}32\\+\underline{12}\end{array}\qquad\begin{array}{r}21\\+\underline{32}\end{array}$$

they can identify those that will give the same sum even without adding. This corresponds to the fact that in operating with sets, $A \cup B$ is postulated to be equal to $B \cup A$. That is, to combine Set B with Set A gives us the same set as combining Set A with Set B. If the sets are finite and disjoint, the corresponding cardinal numbers, a and b, have the property that $a + b = b + a$. This property is named *the commutative law of addition*. Thus, adding numbers is founded on experiences of combining sets of things.

The commutative and associative laws of addition are the principles which make possible checking column addition by adding either from top to bottom or from bottom to top or in some irregular sequence when it is advantageous to forego order to group by tens, and so on. For example, we may add the ones in the example below by adding $6 + 3 + 8 + 4$, or by the order $4 + 8 + 3 + 6$, or $(6 + 4) + 3 + 8$.

$$\begin{array}{r}36\\23\\18\\\underline{44}\end{array}$$

For children it suffices to generalize the associative and commutative laws out of experiences with combining and counting collections or sets. Later, in the secondary school students state these ideas of order

and grouping in addition, as statements about numbers, $a + b = b + a$ and $(a + b) + c = a + (b + c)$.

Subtraction. Children's first experiences in using the idea of subtraction involve the taking of objects from a given collection and determining how many objects remain. Starting with a collection of objects, they separate it into two collections—that which was *left* and that which they *took away*. For example, '$5 - 2 = 3$' where 5 was the number of the set to be separated, describes for them the result of removing a set of objects whose number is 2 and leaving a remaining set of objects whose number is 3.

Subtraction also is used in answering such questions as "How much larger is one group than another?" "How many are taken from a collection of elements if there are so many elements left?" and "How many more are needed to complete a set?" All of these involve subtraction. Furthermore, all of these involve addition in the sense that one of the numbers involved is the sum of the other two.

Thus, addition is a *fundamental* operation, independently defined; subtraction is merely its *inverse* derived from addition and not existing logically until after addition and its algorism have been set up, *e.g.*, if $3 + 4 = 7$, then $7 - 3 = 4$. We emphasize this relationship between addition and subtraction from the outset by teaching addition and subtraction facts together, and by *emphasizing* the presence of the addition idea in the applications made of subtraction. An additional use of this idea occurs when children see that they may check their subtraction problems by addition.

Children learn that the mathematical models, $a + b = c$ and $c - b = a$, may fit several situations in the physical world. They come to realize early in their study of operations with numbers that situations which appear different in the physical world may call for the same mathematical process. Each may help to understand the other. For example, a child who has 5 cents and needs 12 must come to realize that although he must add 7 to 5 to get 12, the number 7 was obtained by subtracting 5 from 12.

Children learn that subtraction is not always possible, and indeed, $4 - 9$ is meaningless in natural numbers. From such experiences elementary school children conclude that they cannot take a larger number from a smaller one. Yet, even while in the elementary school, they may play shuffle board or other games where they may *go in the hole*, and they may hear of temperatures *below zero*. An elementary teacher should not insist that "You cannot take a larger number from a smaller," but should assure her children that such a problem may be worked but that in doing

this, another kind of number is needed. "Later, we will discuss these numbers which make subtraction always possible."

Multiplication. There is a variety of situations which can be used to introduce multiplication of natural numbers. Let us suppose that "each of 5 children at a party is to get 3 balloons as a favor. How many balloons do we need?" "If each of three brothers puts 5 marbles in a bag for his little sister, how many marbles does she get?" "If we have 3 dozen eggs, how many eggs do we have?" Children see these situations as addition problems, addition problems where the groups to be combined are of the same size. Diagrammatically, the balloon problem is illustrated in Figure 13, and the marble problem is illustrated in Figure 13*A*.

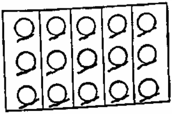


FIG. 13

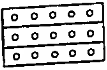


FIG. 13*A*

These diagrams serve to emphasize several important ideas relating to multiplication of natural numbers. Multiplication of natural numbers is equivalent to repeated addition; that is, 5 threes is the same as $3 + 3 + 3 + 3 + 3$. The term 'ab' is equivalent to $b + b + b + \cdots$ for a terms, if a represents a positive integer. In set terms, the cardinal number (2×3) may be defined to be the cardinal number of a set formed by the union of 2 disjoint sets of 3 elements each; i.e., $3 + 3$ or $(a \times b)$ may likewise be defined to be the cardinal number of a set formed by the union of a disjoint sets of b elements each.

The diagram of the marble problem above presents another basic fact. Multiplication of natural numbers is commutative, that is, $3 \times 5 = 5 \times 3$. Children use this idea in learning the multiplication facts. By learning one fact a child has learned two facts, for, although he knows that $3 + 3 + 3 + 3 + 3$ is different from $5 + 5 + 5$, he learns that given the pair of numbers, regardless of the order, their product is the same. Later, this idea is viewed and used as a basic property of all elementary number systems and of many algebras, that is, $ab = ba$, the commutative law of multiplication.

Multiplication of natural numbers is associative; $(2 \times 3) \times 4 = 2 \times (3 \times 4)$ or $6 \times 4 = 2 \times 12$. This idea combined with the commutative law for multiplication leads to such short cuts as one takes when he

mentally views $2 \times 8 \times 5$ as $2 \times 5 \times 8 = (2 \times 5) \times 8 = 10 \times 8 = 80$. In practice adults do not think of all these separate steps. Working with such combinations does help children, however, to build the meanings and understandings which are sometimes labeled 'number sense' and later called 'mathematical maturity'. Similarly when one mentally does 3×20 he should not have to visualize this as

$$\begin{array}{r} 20 \\ \times\ \underline{3} \end{array}$$

and then think $3 \times 0 = 0$ and $3 \times 2 = 6$, but he thinks $3 \times 20 = 3 \times (2 \times 10) = (3 \times 2) \times 10 = 60$. As he gains experience he arrives at a point where he sees 2×40 as $(2 \times 4) \times 10$ or 80, 20×40 as $(2 \times 4) \times 10 \times 10$ as $(2 \times 4) \times 100$ or 800.

The mature arithmetician does not think out each step of this process, nor does he learn it as a process taught in school, but his understanding of the decimal system of numeration and of the associative and commutative principles leads him to make such combinations automatically and subconsciously. This level of learning will be reached only if teachers point out and use these principles repeatedly at all levels of instruction.

Another idea which is used repeatedly at all levels is the *distributive law*. For example, since 5 may be thought of as $3 + 2$ then instead of of 3×5 we may think of it as $3(3 + 2)$ or $3(3) + 3(2) = 9 + 6 = 15$ as illustrated in Figure 14. More generally, for any numbers, a, b, and c, $a \times (b + c) = (a \times b) + (a \times c)$. Since the two sets are equivalent, they have the same cardinal numbers, that is, $(3 \times 3) + (3 \times 2) = 3 \times 5$. We will see in a short time that this idea is essential in deriving a rule for such operations as finding the product of 43 and 65.

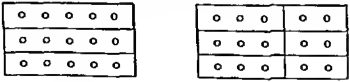


FIG. 14

The commutative, associative, and distributive laws are the basis for deriving most of the important properties of systems of numbers.

Division. We have seen that the model for solving such examples as $3 \times 5 = 15$ is of the form '$a \times b = c$', where the number to be found is c. Should it be necessary to find one of the factors, as $3 \times \square = 15$ or $\square \times 5 = 15$, when given c and one of the factors, either a or b, then the inverse of multiplication is necessary. This inverse process is *division*.

Suppose that we had 15 candies to be tied in packages of 3 each. How many packages of candy would we have? In numbers, $15 \div 3 = n$. We separate the 15 candies into groups of 3, finding that we can make 5 packages. Another problem, conceptually different for elementary school children, is stated by asking what would be the share for each person if we divided 15 candies among 3 children, that is, if we would separate the 15 elements into 3 equal (same number in each) sets. In numbers, $15 \div n = 3$. The first situation involved our taking n groups of 3 from 15 and determining that n was 5. In the second problem we were forming 3 groups from 15 and determined the size of each group to be 5. Both situations were solved by $3\overline{)15}$. These two applications of division are commonly known as quotative (measurement) and partitive (sharing) division situations. Although not so defined in the classroom with children, children are given many experiences with both kinds of problem situations so that they may come to learn to recognize that these different situations have a common solution because they both call for finding one of the factors when we know the other factor and the product.

As with addition and subtraction, an important use is made of the inverse relationship between multiplication and division when multiplication is used to check the result of division. Both the process and its fundamental character should be pointed out to students. If they learn early that any question about division must be settled by reference to the fundamental operation, multiplication, they will not have so much trouble later with the fact that division by zero is impossible. When faced with

$$\frac{0}{6} = ?, \qquad \frac{6}{0} = ?, \qquad \frac{0}{0} = ?$$

they will convert them to $0 = ? \times 6$, $6 = ? \times 0$, $0 = ? \times 0$ and see that the answers are "0," "no number," and "any number" respectively. When the students are more mature they may be warned against the improper use of the word 'infinity' and the symbol '∞' $\left(\frac{6}{0} \neq \infty\right)$, and interested in the ideas associated with the fact that we may write $\lim\limits_{x \to 0} \frac{6}{x} = \infty$ if '∞' has been carefully and properly defined.

OPERATIONS AND THE NUMERATION SYSTEM

In the elementary school one cannot be rigorously logical, but one can build concepts that are consistent and easily extended to more

abstract ideas in the secondary school. At the elementary level we regard the four operations with natural numbers (addition, subtraction, multiplication, and division) as processes of combining two or more sets into one set and separating one set into two or more sets.

The principles of the decimal system of numeration constitute the foundations upon which the rationale of the four operations is built not only with natural numbers but also with decimal and common fractions. For instance, in adding two- and three-place numbers, children learn to analyze situations and to devise algorisms in a variety of ways, as shown in the following example:

$79 + 6$ may be regarded as

$79 + 6 = (70 + 9) + 6$, to be completed by counting, or

$79 + (1 + 5) = (79 + 1) + 5$

$= 80 + 5 = 85$, solved by grouping in tens, or

$79 + 6 = 70 + (6 + 9)$, to be completed by adding ones.

The algorism for adding 32 and 46 may be viewed as

$$\begin{aligned} 32 + 46 &= (30 + 2) + (40 + 6) \\ &= (2 + 6) + (30 + 40) \\ &= 8 + 70 \text{ or } 78, \\ \text{or as, } 32 + 46 &= 32 + 40 + 6 \\ &= 72 + 6 \text{ or } 78. \end{aligned}$$

In doing the above problems children have used the ideas of the commutative and associative laws without ever hearing the words. Later, in algebra, these laws are written as:

$a + b = b + a$ (Commutative law of addition)

$ab = ba$ (Commutative law of multiplication)

$(a + b) + c = a + (b + c)$ (Associative law of addition)

$(ab)c = a(bc)$ (Associative law of multiplication)

We will later note the use of these in factoring, in creating an understanding of the operations with signed numbers, and as a guide in extending our number system.

In subtraction, the place value idea and the process of separation are basic principles underlying both the meaning and the algorism for the

operation. The numeration system indicates the way in which the algorism shall be written. It also is the basis on which we operate in the process of regrouping a larger unit into smaller units, such as,

$$\begin{array}{r} 7000 \\ -\ 329 \\ \hline 6671 \end{array}$$

where 700 tens is regrouped into 699 tens and 10 ones, as

$$\begin{array}{r} 6990 + 10 \\ -\ 320 - \ \ 9 \\ \hline 6670 + \ \ 1 \end{array} \qquad \text{or} \qquad \begin{array}{rr} 699 & 10 \\ -\ 32 & 9 \\ \hline 667 & 1. \end{array}$$

In multiplication, the system of numeration and the distributive law give us a way of writing the partial products. For example, 46×23 is written $(40 + 6)(20 + 3)$ or $(40 \times 20) + (40 \times 3) + (6 \times 20) + (6 \times 3)$. This is commonly written as

$$\begin{array}{r} 23 \\ \times 46 \\ \hline \end{array}$$

where we find 6 twenty-threes or $(6 \times 3) + (6 \times 20) = 138$. Similarly, 40 twenty-threes is written $(40 \times 3) + (40 \times 20) = 920$. We may write

$$\begin{array}{r} 23 \\ \times\ 46 \\ \hline 138 \\ 92\ \\ \hline 1058 \end{array}$$

where positional notation makes it possible to express 920 by '92'.

The division process too depends on our system of numeration. If we think of $465 \div 3$ as separating 465 into 3 equal groups then we think of our problem as $3\overline{)400 + 60 + 5}$ or as $3\overline{)300 + 150 + 15} = 100 + 50 + 5 = 155$. However, if we think of $465 \div 3$ as finding the number of groups of 3 in 465, then this may be viewed as subtracting 3 in multiples of 100, 10, and so on, as

$$\begin{array}{r|r} 3\overline{)465} & \\ \underline{300} & 100 \\ 165 & \\ \underline{150} & 50 \\ 15 & \\ \underline{15} & \underline{\ \ 5} \\ & 155 \end{array}$$

Just as the place-value principles of our decimal system of numeration and the general principles for operations with numbers have been used here in the development of the algorisms we now employ, so we should take advantage of every opportunity to illustrate and emphasize the meaning, use, and importance of these principles. When we attack our mathematical problems by using automatically the ways of saying, thinking, and writing numerals we have been taught, we are manipulating by habit, without reference to meaning. While we possess a perfect symbolism, we often use it with little sense. The *system* thinks for us! It is good to set up habits of accurate, rapid computation, but computation will be quicker and more accurate and the processes will be remembered longer and more easily applied to new situations if they have been learned through developmental processes with an understanding of their meaning.

SPECIAL NUMBERS

One. Every natural number is *special*, just as is every person, because it is unique. However, 1 is *extra special* for several reasons: (1) it is the first natural number, (2) it was the child's building block out of which all natural numbers may be formed (any number plus one produced a *successor* number), and (3) 1 times a number equals that same number ($1 \times n = n$). It is for this last reason that mathematicians call 1 the 'multiplicative identity element'. It is the element in the set of natural numbers which when used as a multiplier with any other element produces that other element as the product. For example, 1 four or $1 \times 4 = 4$. $1 \times a = a$. This property is retained by the natural number 1 as the number system is extended through elementary mathematics into algebra and into more advanced mathematical topics.

This property is the basis for several other properties. It is from this property that we know that $a \div 1 = a$, because $a = 1 \times a$. Similarly $a \div a = a/a = 1$ $(a \neq 0)$ because $a = 1 \times a$. That $a/1 = a$ and $a/a = 1$, for every a except 0, are important ideas in dealing with fractions. For example,

$$\frac{24}{12} = \frac{12}{12} \times \frac{2}{1} = 1 \times 2 \text{ or } 2.$$

Stressing these facts early helps to avoid later difficulty in adding *mixed numbers* in both arithmetic and algebra.

Zero. Zero is the cardinal number assigned to the *empty* or *null* set, the set containing no elements. This is familiar in everyday experiences—the empty cookie jar, the *no score* made in the dart game when the dart always went off the board, no money left from the allowance, no pencil.

In some respects zero is very much like *one*. Zero is an *identity* element. It is the additive identity, that is, $a + 0 = a$, for all numbers a. As with 1, this leads to additional properties, namely, $a - 0 = a$ and $a - a = 0$. Zero also presents special properties and problems in connection with multiplication and division. We will discuss these uses, properties, and problems later.

Zero and *one* are most important numbers.

RATIO-LIKE NUMBERS

The natural numbers are adequate for counting and calculating when we wish to determine *how many*. But how can we use these numbers to measure a half cup of milk, a part of a yard of ribbon, one of three equal shares of a quantity of meat? For these purposes, the natural numbers are inadequate. We need new numbers. Fractions were invented to deal with parts of things and to make division always possible. We recall that given a pair of natural numbers such as 5, 3 there was not always a third number to represent the quotient, $5 \div 3$.

Fractions. Historically, fractions owe their creation to the transition from counting to measuring. The very word 'fraction' tells us something about the early significance of these numbers for it comes from the same Latin root as *fracture*; in German, fractions were once called *gebrochen zählen* or *broken numbers*. The symbol for a fraction is a pair of numerals. For the elementary school student one member of the pair specifies the number of parts into which the unit has been divided and the other of the pair specifies the number of these parts with which we are concerned.

The earliest occurrence of fractions dates back almost to the prehistoric period. In the earliest Egyptian writings, we find that the Egyptians used fractions which today we describe as unit fractions, e.g., '$\frac{1}{2}$', '$\frac{1}{3}$,' '$\frac{1}{4}$'. Instead of restricting themselves to fractions with unit numerators, the Babylonians restricted their denominators to 60 and the powers of 60. It is to this fact that we attribute our modern use of a minute as $\frac{1}{60}$ of a degree or of an hour, and a second as $\frac{1}{60}$ of a minute. This restriction of the denominator to powers of one number is basic to our 'decimal fractions' in which denominators of 10 and powers of 10 are indicated by the position in which we write the digits 0, 1, 2, . . . 8, 9.

The first recognition that *any* pair of numbers might be used to represent a new number, a fraction, occurred in the early days of Greek mathematics. So, we find the development of fractions from those which were limited in their numerator or in their denominator to those which could have any number as a numerator and any number, excluding zero, as a denominator.

Today, in modern mathematics, fractions may be conceived of as an ordered pair of integers. This concept of fractions is more abstract than the notion of fractions as associated with things such as measures of grain, pieces of pie, or plots of land. Before discussing this mathematical approach, let's turn to some daily life applications of fractions used in the initial development of these ideas in the elementary school.

Interpretations of Rational Numbers. A fraction or rational number such as $3/5$ may be used in a variety of ways to describe physical situations. In some uses fractions are measures of a quantity or the result of an operation; in others they represent relations between sets.

As a *measure*, the fractional numeral, '$3/5$', may be conceived of as (1) naming the number property of a set of 3 elements each of which is $1/5$ of some unit. If Figure 15 represents a candy bar divided into 5 equal parts, we may represent $3/5$ as the 3 shaded parts. (2) It may be regarded as representing $1/5$ of 3 units. If Figure 16 represents 3 candy bars, placed side by side, and divided into five equal parts, the shaded portion represents $1/5$ of the 3 candy bars. Also $3/5$ may be regarded as representing the outcome of the process of division. For example, $3 \div 5 = 3/5$, because, as we will see later, $5 \times 3/5 = 3$. Further, when 13 is divided by 5, the quotient earlier expressed as 2 with a remainder of 3 may now be written as $1^3/5$ or $2 + 3/5$.

FIG. 15

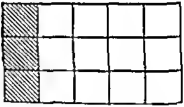


FIG. 16

As a representative of a *relation*, the fractional numeral designates a ratio. The ratio $3/5$, which also may be expressed as "the ratio of 3 to 5," or "3 out of 5," or "3:5," is interpreted in its applications to mean that given 2 groups of objects, for every 3 elements in one group, there are 5 elements in the other. The two groups may contain 9 and 15 objects respectively, or 30 and 50, 45 and 75, and so on. The essential condition is that if the first group contains $3 \times k$ objects, then the second group contains $5 \times k$ objects. Similarly if the ratio of the side of a square to its diagonal is given as 1 to $\sqrt{2}$, this does not mean that the particular square being considered has a length of 1 unit and its diagonal is $\sqrt{2}$ units but rather these 2 lengths may be represented by 2 numbers such that if the first is some number, say n, units long, then the length of the second is represented by the number, $(\sqrt{2} \times n)$.

Only *integers* such as 1, 2, 3, 4, ... (and later the negative integers $-1, -2, -3, -4, \ldots$) may be used to make *rational fractions* such as $\frac{3}{5}$. However, our last example, the ratio of 1 to $\sqrt{2}$, indicates that we may also use any real numbers such as $\sqrt{2}$, π, $\sqrt[3]{5}$ in forming ratios. We will discuss such numbers later. Here we will illustrate ratio in terms of pairs of positive integers. If a and b are a pair of numbers which are used to name a ratio a/b or (a, b), then all pairs of numbers (ka, kb) obtained by multiplying a and b by the same number k represent the same ratio. Thus the ratio 3/5 may also be represented by the number pairs (9, 15), (12, 20), (15, 25), (30, 50), (36, 60), (45, 75), and so on, all of which represent the same ratio by this definition. In fact, a ratio can be defined as a set of ordered number pairs all of which are equal in accord with this or a similar definition of equality.

A typical ratio problem would then be: "If the ratio of the number of Pete's hits to the number of times he has been at bat is 300 to 1000, and he has been at bat 20 times, how many hits has he made?" This merely says the (300, 1000) and (?, 20) are both members of the same set of equal number pairs. Noting that $\frac{1}{50} \times 1000 = 20$ we realize that the number represented by "?" must be $\frac{1}{50} \times 300 = 6$. This problem also illustrates that a ratio may frequently be thought of as a *rate*. Thus Pete was hitting at the rate of 300 hits per 1000 times at bat or of 3 hits per 10 times at bat or of .3 hits per time at bat. Of course, .3 hits is nonsense in baseball, but the idea of a ratio gives meaning to the statement as a whole.

We recognize, of course, that this problem could have been written and solved as an *equation*, $1000n = 300 \times 20$. This is often called a

$$\frac{300}{1000} = \frac{n}{20},$$

proportion, or as the substitution of 20 for b in another equation, sometimes called a *formula*,

$$h = \frac{300}{1000} \times b,$$

where b represents the number of times Pete has been at bat and h represents the number of hits he will have had if he is *batting 300*. We believe that all of these approaches, and two others, should be taught clearly and as related to one another as the topic of ratio is developed through the grades at progressively higher levels of maturity. The two other approaches we refer to are the graphical approach and the approach via the power function $y = ax^n$ which ultimately unifies direct

and inverse variation with graphs, ratio, and proportion. This power function will be discussed in the next chapter, "Relations and Functions," which will also extend further the discussion of sets of ordered pairs of numbers.

Let us return here to the discussion of rational numbers and their equivalence which we began above. We recall that the early Greeks were the first to use *any two* integers in writing fractions. We now note that some Greek writers wrote the two numerals on one line thus, 3,5 (using Greek numerals); some wrote the denominator above the numerator, $\frac{5}{3}$; some wrote the numerator above the denominator, $\frac{3}{5}$. In all these cases there were *three* central ideas: (1) new numbers, today called fractions, rational numbers, or rational fractions, were constructed out of *old* numbers; (2) a rational fraction is a pair of natural numbers; (3) the two numbers of a particular pair do not both play the same role, that is, they are not interchangeable and must be distinguished from each other.

Early Ideas of Equivalence Classes. Just as a natural number can be designated by such different numerals as 5, $(3 + 2)$, $(4 + 1)$, so a rational number can be designated by different pairs of numerals. As a common concrete approach, a unit, as in Figure 17, which has been divided into 2 equal parts may be divided into 4 equal parts by dividing each half into two equal parts or into 8 equal parts by dividing each

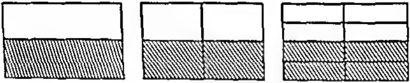


FIG. 17

half into 4 equal parts. From this we see intuitively that two-fourths is equal to one-half and that four-eighths is equal to two-fourths or to one-half. From this and other examples students discover that a new but equivalent fraction may be derived from another fraction by multiplying its numerator and denominator by the same number. We then may have an infinite set of different number names for the number property of a part of a unit. From such experiences evolves the generalization that if the numerator and denominator of a fraction are multiplied by the same number or divided by the same number, the resulting new fraction is equal to the earlier fraction. In a purely logical organization of the fractional part of our real number system, we will see that this becomes a theorem derived not from diagrams or objects but from a definition of equal fractions.

FRACTIONS AS ORDERED PAIRS OF NATURAL NUMBERS

Probably the best definition of $3/5$ from a mathematical viewpoint is that $3/5$ is merely an ordered pair of natural numbers which could just as well have been written (3, 5). The *ordered* merely means that (3, 5) is not the same as (5, 3) just as $3/5$ is not the same as $5/3$. The order in which the numbers are named makes the difference. We will explore this mathematical definition further in the next section. The preceding concrete interpretations may be used as steps leading students over the years from representations of fractions by things to the abstraction of a rational number. The ordered pair (3, 5) together with the mathematical principles governing the operations with it are a single *mathematical model* which can be used to represent these several types of conceptual, physical, or concrete situations as 3 of 5 equal parts of a unit, or a comparison between a collection of 3 and a collection of 5.

A rational number is also on occasion defined as a number which may be written as p/q, the quotient or ratio of two integers p and q. Our definition that a rational number is an ordered pair of integers (p, q) is equivalent to this. Although logically arbitrary, this abstract approach to our definitions is being guided by our intuitive identification of the pairs (1, 2), (2, 3), (3, 2), (2, 1) with our old friends the rational fractions $1/2$, $2/3$, $3/2$, $2/1$.

We now ask, "Are two different pairs ever *equal* and if so when?" This question cannot be answered with logical rigor until we define equality for ordered pairs of integers. Past experience with fractions suggests that a useful definition is: *Two ordered pairs of integers (p, q) and (a, b) are equal if and only if $pb = qa$.* This defines equality of our new ordered pairs in terms of the equality of our old friends, the integers, since p, q, a, and b are integers. (Actually, in this chapter, we have not yet extended our number system from the *natural numbers* to the *integers*, but all that is said here applies to both.) For example, $(1, 2) = (2, 4) = (3, 6)$ and $(3, 5) = (6, 10)$ but $(1, 3) \neq (3, 1)$ nor does $(2, 0) \neq (0, 2)$. These statements follow from the facts that $1 \times 4 = 2 \times 2$; that $2 \times 6 = 4 \times 3$; that $3 \times 10 = 5 \times 6$; that $6 \times 15 = 9 \times 10$; that $1 \times 1 \neq 3 \times 3$; and that $2 \times 2 \neq 0 \times 0$. These equalities furnish the real mathematical proof that $1/2 = 2/4 = 3/6$. We derive the equal pairs from one another by multiplying (or dividing) each integer of one pair by the same integer. In general terms, the theorem that $(p, q) = (kp, kq)$ follows from our definition of equality of pairs, because $p(kq) = q(kp)$. This is the mathematical proof of the rule that a fraction is unchanged if its numerator and denominator are multiplied or divided by the same number. We previously *discovered* this by breaking into parts such materials as geometric figures, pieces of pie, and candy bars.

OPERATIONS WITH RATIONAL NUMBERS

Earlier in this chapter we saw how we could operate with the natural numbers. We recognized the usefulness and importance of the fact that the natural numbers obey associative, commutative, and distributive laws with respect to the operations of addition and multiplication. We would hope to define operations with fractions such that these laws still hold true.

Addition and Subtraction. We recall '$2 + 3$' expressed the number resulting from the union of two sets whose cardinal numbers were 2 and 3 and that we could also express the result of this union by the numeral, '5'. At the early intuitive level we associated '$1/2$' with 1 part which has been obtained by dividing a whole into 2 equal parts. If we wish to combine $1/2$ of a unit and $1/3$ of that same unit, the sum may be written as '$1/2 + 1/3$'. However, if we wish to express this as one fractional numeral, then we must get a common unit of measure. Earlier we have noted that fractions may be changed to equivalent fractions if we multiply their numerators and the denominators by the same nonzero number. Our problem then becomes that of finding a common unit that can be used as a measure for each of the two fractions. We regard $3/6$ as $3 \times 1/6$ and $2/6$ as $2 \times 1/6$. Thus 3 one-sixths and 2 one-sixths are 5 one-sixths. This is consistent with addition of natural numbers for just as $3 + 2 = 3 + 2$, $3/6 + 2/6 = 2/6 + 3/6$. Similarly, just as $(1 + 3) + 2 = 1 + (3 + 2)$, or $4 + 2 = 1 + 5$, $(1/6 + 3/6) + 2/6 = 1/6 + (3/6 + 2/6)$ or $4/6 + 2/6 = 1/6 + 5/6$. We may write '$1/2 + 1/3$' as '$3/6 + 2/6$' or '$5/6$'. Thus, addition of fractions involves first getting equivalent fractions with the same number as the denominator and then adding the numerators.

Another important fact which follows from the definition of equality of fractions is that $2/2 = 3/3 = 67/67 = 100/100 = a/a = 1$. This should be associated with several types of problems and related ideas. For example, in dealing with improper fractions $4/3$ may be regarded as $\dfrac{3+1}{3}$. This becomes $\dfrac{3}{3} + \dfrac{1}{3} = 1 + \dfrac{1}{3} = 1\dfrac{1}{3}$. Let us note the relationship of the fact, $a/a = 1$, to the processes of addition and subtraction.

$$\begin{array}{l} 3\frac{7}{8} \\ \underline{2\frac{3}{8}} \\ 5\frac{10}{8} = 5 + \frac{8}{8} + \frac{2}{8} = 5 + 1 + \frac{1}{4} = 6\frac{1}{4}, \end{array}$$

$$\begin{array}{lcr} 4\frac{3}{5} = 3 + \frac{5}{5} + \frac{3}{5} & = & 3\frac{8}{5} \\ \underline{-1\frac{4}{5}} & = & \underline{-1\frac{4}{5}} \\ & & 2\frac{4}{5} \end{array}$$

Children come to see that the idea in the relationship $a/a = 1$ is related to the regrouping necessary to combine fractions in addition and subtraction.

We have emphasized that in dealing with natural numbers subtraction is the inverse of addition. We have defined the operations with fractions such that they have the same basic properties. The commutative and associative laws hold for addition of fractions just as they did with natural numbers. The process of subtracting fractions is the inverse of adding fractions; that is, $5/6 - 3/6 = n$ such that $3/6 + n = 5/6$, or $n = 2/6$.

Multiplication and Division. In developing an understanding of the meaning of fractions, we learned that fractions not only enable us to deal with parts of things but that they also make possible the division of one number by another. We recall that multiplication of natural numbers could be regarded as repeated addition; that is, 3 fives (3×5) could also be obtained by adding five and five and five. This definition of multiplication soon breaks down if the number system is extended to include numbers other than the natural numbers. However, we use this idea to help us find a suitable definition for the multiplication of fractions. For example, 3 one-fifths or $3 \times 1/5$ can be viewed as $1/5 + 1/5 + 1/5$ or as $3/5$. If we wish our new numbers, fractions, to behave as did our old numbers, then it must be true that $1/5 \times 3$ will have the same result as $3 \times 1/5$. We verify this at the elementary level with diagrams as in Figure 18.

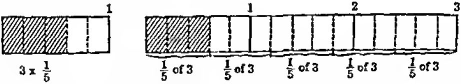


FIG 18

In using this last diagram (Fig. 18) it should be pointed out not only that the shaded portion is "$1/5$ of 3" by reference to earlier discussions, but also that it represents the answer to $3 \div 5 = n$, *because* $3 = 5 \times n$. We then make the generalization that the product of a fraction and a whole number is a new fraction whose numerator is the product of the whole number and the numerator of the fraction and whose denominator is the denominator of the fraction.

In summary, when the multiplier is a fraction, the idea of repeated addition gives us no direct help, but the commutative law makes it applicable. In the minds of elementary school children, $1/5 \times 3 = 3/5$ is similar to an earlier experience "$1/5$ of 3" $= 3/5$. This dictates that if

a physical object such as a length of 3 inches were to be divided into 5 equal lengths, each would be $\frac{3}{5}$ of an inch. We may either divide 3 by 5 or we may multiply 3 by $\frac{1}{5}$. These are the early experiences which ultimately lead to the generalization that dividing by a number gives the same results as multiplying by its reciprocal. All of these ideas and approaches at some time fit somewhere into a growing and maturing idea of fractions.

However, because its reasoning is somewhat sophisticated for elementary school children, an initial approach to learning to multiply $\frac{1}{5} \times 3$ from $3 \times \frac{1}{5}$ by the idea of commutativity is not too desirable. There also is a danger that students may think they are deriving the result rather than just defining it. Consequently using an inductive approach has merit both psychologically and logically as a way of developing their beginning ideas. In using it we may start writing

$$\begin{array}{lcl} 4 \times 12 = 48 & \text{and} & 3 \times 12 = 36 \\ 2 \times 12 = 24 & & 1 \times 12 = 12 \\ 1 \times 12 = 12 & & \frac{1}{3} \times 12 = 4. \\ \frac{1}{2} \times 12 = 6 & & \\ \frac{1}{4} \times 12 = 3 & & \end{array}$$

Also,

$$\begin{array}{lcl} 6 \times \frac{1}{2} = 3 & \text{and} & 6 \times \frac{1}{3} = 2 \\ 3 \times \frac{1}{2} = \frac{3}{2} & & 3 \times \frac{1}{3} = 1 \\ 1 \times \frac{1}{2} = \frac{1}{2} & & 1 \times \frac{1}{3} = \frac{1}{3} \\ \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} & & \frac{1}{2} \times \frac{1}{3} = \frac{1}{6} \\ \frac{1}{4} \times \frac{1}{2} = \frac{1}{8} & & \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}. \end{array}$$

By so doing, children come to realize that as the multiplier increases, the product does likewise, and that if the multiplier is divided by 2 or by 3, the product is likewise divided by 2 or 3.

This inductive approach makes three important additional points. First, when a multiplier is less than 1, the product will be less than the other factor. Later, children see that if both factors are less than 1, the product is less than either in contrast to that which had been true in multiplication of natural numbers. Secondly, it focuses attention on a general pattern. It is a means of helping children look for systematic changes and relationships, to think in terms of *if* and *then*. For example, if in multiplication we decreased one factor, we decrease the product; or more precisely, if we divide one factor by any number, then the product will be divided by the same number. Finally, and still more generally, every time a student discovers something for himself by any process,

he is getting some introduction to the processes of problem solving and creativity.

Just as we related the multiplication of fractions to that of natural numbers, we relate the division of fractions to that of natural numbers. We may now ask the question, "How many one-thirds are there in 2?" We recall from our early ideas of fractions that there are three $\frac{1}{3}$'s in 1. By this children are led to understand that $1 \div \frac{1}{3} = 3$, and from this that $2 \div \frac{1}{3} = 6$. Further examples will lead to the generalization that division of a number by $1/n$ is equivalent to multiplication by n. From Figures 19 and 20, they can also see the idea of division by regrouping. (1) How many $\frac{3}{4}$'s in 6? That is if $6 \div \frac{3}{4} = n$, then $n = \frac{24}{4} \div \frac{3}{4} = \frac{24}{3} = 8$. Or (2) How many $2\frac{1}{2}$'s in $8\frac{1}{2}$? If $8\frac{1}{2} \div 2\frac{1}{2} = n$, then $\frac{17}{2} \div \frac{5}{2} = \frac{17}{5} = 3\frac{2}{5}$. Thus, $n = 3\frac{2}{5}$.

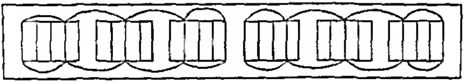


FIG. 19

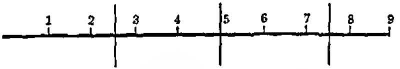


FIG. 20

We can also develop processes for such problems as $1 \div \frac{2}{3}$ by noting that since $\frac{2}{3}$ is twice as much as $\frac{1}{3}$, or $2 \times \frac{1}{3} = \frac{2}{3}$, then we would expect $1 \div \frac{2}{3}$ to be only half as much as $1 \div \frac{1}{3}$. Since $1 \div \frac{1}{3} = 3$, then $1 \div \frac{2}{3} = \frac{3}{2}$.

Or, $1 \div \frac{2}{3} = (1 \times 3) \div 2 = 1 \times \frac{3}{2}$.

This inductive process may then be continued with such examples as

$$4 \div \tfrac{1}{3} = 4 \times 3 = 12$$

$$4 \div \tfrac{2}{3} = (4 \times 3) \div 2 = 6, \text{ or } 4 \div \tfrac{2}{3} = \frac{4 \times 3}{2} = 6,$$

which may be generalized to

$$\frac{4}{5} \div \frac{2}{3} = \frac{4}{5} \times \frac{3}{2} = \frac{4 \times 3}{5 \times 2}$$

or to the familiar statement of the short cut—*to divide by a fraction, invert the divisor and multiply.*

As a student begins his study of algebra, he is ready for more emphasis on the inverse relationship between multiplication and division. He sees that

$$\frac{a}{b} \div \frac{c}{d} = \frac{?}{?} = \frac{x}{y}$$

is equivalent to

$$\frac{a}{b} = \frac{x}{y} \times \frac{c}{d}$$

We see from this that the numbers to be substituted for x and y must be such that $x/y \times c/d = a/b$. At first in numerical examples such as $3/4 \div 2/5 = x/y$, we work out step by step the replacement of x/y by numbers such that $3/4 = x/y \times 2/5$ or $3/4 = (\ ) \times 2/5$. We think: "We could achieve our goal of converting $2/5$ to $3/4$ in two steps. First we could think of a fractional numeral which if put in the parentheses would make the right-hand member a symbol for one. This first substitution would then be $5/2$, since $5/2 \cdot 2/5 = 1$. We must next put into the parentheses the number symbol which will convert this from 1 to a symbol for $3/4$. This must then be $3/4$, or $3/4 = (3/4 \times 5/2) \times 2/5$ or $x/y = 3/4 \times 5/2$." From experiences as these, students see that if $a/b \div c/d = x/y$, then $x/y = a/b \times d/c$.

A second approach whereby a student may increase his understanding of division of fractions is by writing

$$\frac{a}{b} \div \frac{c}{d} \qquad \text{as} \qquad \frac{\frac{a}{b}}{\frac{c}{d}}$$

Then the teacher leads him to review the ideas that fractions with unit denominators are simpler, and that if one multiplies the numerator and denominator by the same number, he derives an equivalent fraction. The following gives some idea of the way a student of algebra might think to arrive at the algorism.

$$\frac{a}{b} \div \frac{c}{d} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{\frac{a}{b} \cdot \frac{d}{c}}{\frac{c}{d} \cdot \frac{d}{c}} = \frac{\frac{a}{b} \cdot \frac{d}{c}}{1} = \frac{a}{b} \cdot \frac{d}{c}.$$

Note the recurring emphasis on $\frac{a}{a} = 1$, and $(a \div a) = a/a = 1$.

DECIMAL FRACTIONS AND PER CENT

A child's first experiences with decimal fractions are as a set of fractions whose denominators are limited to 10, 100, 1000, $\cdots$ 10^n, and which can therefore be represented by use of the place value concept in a number system based on 10. We derive the procedures for converting so-called common fractions to decimal fractions, and conversely, from our earlier notions of rational fractions and our definition of decimal fractions. Thus the problem of converting $2/25$ to a decimal fraction merely asks $2/25 = ?/100$. We find the fourth number by determining by what we must multiply 25 in order to obtain 100 and then multiplying 2 by that same number, that is,

$$\frac{2 \cdot x}{25 \cdot x} = \frac{n}{100}.$$

Thus $n = 2 \times 4 = 8$, and $2/25 = 8/100 = .08$. At more mature levels, these steps are seen to merge into one; we merely divide 2 by 25. This one-step process may also be approached directly as based on the fundamental idea that a fraction is the quotient of two numbers.

Children first meet never-ending or infinite processes when they find that $1/3 = .333 \cdots$. Rather than avoiding or slighting discussions of infinite processes and the limits which are associated with them, their abstractness and importance should be motivated by some attention to them early and by then returning to them in later work. *Understandings are not all-or-nothing affairs, but grow, develop, and expand as a concept is met repeatedly in different contexts and at different levels of abstraction and generality.* With this in view it becomes important that we point out and reteach such matters as often as is possible with many examples at successively higher levels of organization, and with the recognition that we aren't expecting complete understanding and mastery at the first approach, but that we are building toward such a goal. Thus the remarkable fact that $1/3$, $1/6$, $1/7$, and so on, lead to infinite, repeating decimals is first merely observed; it may then be illustrated by showing that $1/3$ is less than .4 and greater than .3. An enlarged scale would show $.33 < 1/3 < .34$; $.333 < 1/3 < .334$; and so on. Intuitively we see that $1/3$ corresponds to a single point which is within the nested intervals .3–.4, .33–.34, .333–.334, and so on, and we gain some feeling for $1/3$ as the limit of $.333 \cdots$ but as not exactly equal to any terminating decimal fraction.

Two questions may arise quite naturally from these situations. They are "How can we tell which fractions will terminate and which will not?", and "Can we reverse the process? Is there an ordinary fraction

associated with every infinite decimal?" Each of these questions can be answered at different levels of rigor and abstraction and by using different mathematical tools. We will only sketch some of these answers here.

Every fraction whose denominator is the product of 2's and/or 5's will produce a terminating decimal fraction. (Such a fraction is sometimes referred to as an infinite decimal fraction which after some point repeats only zeros.) We can see this because if we have any fraction such as

$$\frac{7}{200} = \frac{7}{2 \times 2 \times 2 \times 5 \times 5}$$

we could make its denominator a power of 10 by pairing each 2 with a 5 *as long as they last* and then supplying any 5's or 2's needed to complete pairs by multiplying the numerator and denominator by the same factors. Thus, $\frac{7}{200} = \frac{7 \times 5}{(2 \times 5) \times (2 \times 5) \times (2 \times 5)} = \frac{35}{1000} = .035$. Conversely, if the denominator contains any factor which is not a factor of 10 it will be impossible to make it into a fraction whose denominator is a power of 10 and hence, it can not be written as a terminating decimal fraction. Thus, whether or not a fraction terminates depends on the relationship of the factors of its denominator to the factors of *10*, the base of our number system, after the fraction is written in its lowest terms.

We can show that a unique rational fraction corresponds to every repeating decimal. Perhaps the most elementary approach is typified by the following example. Assume that we have emphasized that the dots following $.\overline{142857}$ .. mean that it is really an infinite decimal and the bar over $.\overline{142857}$ means it repeats those same digits in the same order, thus, .142857 142857 142857 .... Then, if N is the rational number represented by $.\overline{142857}$ ..., $1{,}000{,}000\ N = 142857.\overline{142857}$ .... Subtracting one N from $1{,}000{,}000\ N$, we have $999999\ N = 142857$ or $N = \frac{142857}{999999} = \frac{1}{7}$.

When geometric sequences and series are studied later, this problem can be solved in a second, more mature, and more rigorous manner. A geometric series is the sum of a sequence of terms such that each successive term is obtained by multiplying the preceding term by a constant, called the common ratio. Thus the following are geometric progressions:

$$2 + 6 + 18 + 54 + 162$$
$$\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} + \tfrac{1}{32} + \cdots$$
$$a + ar + ar^2 + ar^3 + \cdots ar^{n-1} + ar^n + \cdots.$$

The sum of n terms of the latter is $S_n = \frac{a - ar^n}{1 - r}$, and, if r, the common ratio, is less than 1 in absolute value, the limit of S_n as n increases without bound exists and is given by

$$\lim_{n\to\infty} S_n = \frac{a}{1 - r}.$$

This limit is called the sum of the progression.

Periodic decimal fractions then are but geometrical series in disguise. For example, the infinite decimal fraction $0.\overline{21}\ldots$ or $0.212121\ldots$ actually means

$$\frac{21}{100} + \frac{21}{10{,}000} + \frac{21}{1{,}000{,}000}\cdots.$$

This is a geometrical series with first term, $a = \frac{21}{100}$ and common ratio, $r = \frac{1}{100}$. The summation formula shows that

$$\lim_{n\to\infty} S_n = \frac{\frac{21}{100}}{1 - \frac{1}{100}} = \frac{21}{99} = \frac{7}{33} = .2121\overline{21}\cdots.$$

When infinite repeating decimals are discussed we should also refer to and illustrate the fact that there are also numbers which are represented by infinite *non*repeating decimals. These are the irrational numbers such as $\sqrt{2} = 1.4142\ldots$ and $\pi = 3.14159\ldots$ which are represented by infinite decimals that never repeat the same digits in the same order.

In order to compute with these numbers, we must cut off the infinite decimal at some point. When we do this, we select a rational number to be used in computation. Rounding off these infinite decimals introduces approximations in our work. It is here that students should come to understand that there are mathematical theories of approximation and approximate computation, which are good, important, and correct branches of mathematics and not mere substitutes for careful work or lack of knowledge. Computation with approximate numbers will be discussed in a later chapter of this book.

Problems Involving Decimal Fractions and Per Cents. Problems involving decimal fractions and per cents most frequently involve three numbers, as 25% of 48 is 12, or .2 of 15 is 3. In practical problems any

one of these three numbers might be unknown at the outset. Thus our first example might have arisen out of any one of the three questions:

What is 25% of 48?
What per cent of 48 is 12?
12 is 25% of what number?

Because these three situations typify practically all per cent problems as well as those which occur in connection with decimals, common fractions, and ratio and proportion, teachers and texts have often taught problem solving under the heading of three *cases* with three associated *rules*. We believe that this formalizing of the problem solving process is undesirable from the viewpoints of meaningful teaching and continuity. The unifying idea suggested above is merely that in every case we are given a situation in which the product of two numbers equals a third [rate $\times$ time $=$ distance; per cent (in hundredths) $\times$ base $=$ percentage; principal $\times$ rate $=$ interest (or discount); sales $\times$ rate $=$ commission, and so on]. Students begin to solve problems of this type as soon as they begin to learn multiplication. (As soon as they learn $2 \times 3 = 6$, they should begin to think "if $2 \times$ something $= 6$, the 'something' is 3.") They continue to solve problems of this type as they learn that division is the inverse of multiplication. They progress through such stages as are represented by: "$\frac{1}{4}$ of $12 = n$"; "$\frac{1}{4} \times 12 = n$"; "$\frac{1}{4} = n/12$"; "$.25 \times 12 = n$"; "25% of $12 = n$." Finally, in algebra, they represent all problems of this type by the general sentence $x \cdot y = z$ and acquire the ability to solve for one unknown whenever they are given values of the other two. If the emphasis throughout has been on the underlying common relationship and the connection between multiplication and division, we believe they will not only be able to handle whatever *case* comes along, but will also have begun to understand and appreciate how one *mathematical model* may represent many different daily life and physical situations.

RATIONAL NUMBERS AS ORDERED PAIRS

We have said that our *new* rational numbers may be constructed as pairs of old numbers, and we defined two pairs (p, q) and (a, b) to be equal if and only if $pb = qa$. Now let us consider the fundamental operations from this abstract viewpoint.

Addition and Subtraction. With the idea that a fraction is merely an ordered pair of integers, we have no need to talk of *like quantities* or common denominators. Instead, we *define* addition of rational numbers as $(a, b) + (c, d) = (ad + bc, bd)$. The ordered pair formulation

f our problem $\frac{1}{2} + \frac{1}{5}$ would be $(1, 2) + (1, 5) = [(1)(5) + (2)(1),$ $2)(5)] = (7, 10)$ or $\frac{7}{10}$. From this definition of addition and our earlier efinition of equality we can show that addition of fractions also obeys ne commutative and associative laws, so important for the natural umbers. For example, by definition $(a, b) + (c, d) = (ad + bc, bd)$ and $c, d) + (a, b) = (cb + da, db)$. The two right hand pairs are the same, owever, because by the commutative law for natural numbers, $cb =$ $c, da = ad, db = bd$. Thus, $(ad + bc, bd) = (cb + da, db)$. We will leave t to our reader to show that addition of these pairs is associative, but ve would like to point out how to derive a rule for subtraction of pairs rom the rule for addition.

Subtraction is the inverse operation of addition. In a subtraction problem we are given the result of an addition and one of the addends and asked to find the other addend. Starting with $\frac{3}{4} - \frac{1}{3} = ?/?$ and converting this to the ordered pair form, it becomes $(3, 4) - (1, 3) =$ $(?, ?) = (x, y)$. Converting this to addition, the fundamental operation, we have $(3, 4) = (x, y) + (1, 3)$. From this we derive $(3, 4) = (3x +$ $y, 3y)$. Our definition of equality states this will be true if and only if $3(3y) = 4(3x + y)$ or $5y = 12x$. This last relationship is satisfied if $x = 5$ and $y = 12$, that is, if $(x, y) = (5, 12)$ i.e. $\frac{3}{4} - \frac{1}{3} = \frac{5}{12}$. Note that $5y = 12x$ if $x = 10$ and $y = 24$, or $x = 15$, $y = 36$, and so on. These correspond to the pairs (10, 24), (15, 36), all of which are equal. This illustrates the fact that any single fraction represents a family or set containing an infinite number of different equal fractions. Such a set is sometimes called an *equivalence class*. When this terminology is used, a fraction may be defined as the equivalence class of equal pairs of integers. From this viewpoint no particular pair is regarded as the fraction itself, but each pair is a representative of the equivalence class which is the fraction. This idea of a fraction parallels the conception of a cardinal number as the common property of all sets which can be put into one-to-one correspondence with each other. Any one of the sets could be used as a representative set.

Our arbitrary definition, $(a, b) + (c, d) = (ad + bc, bd)$, corresponds to what we get if we apply the rules of ordinary ninth grade algebra to $\frac{a}{b} + \frac{c}{d}$, namely $\frac{ad + bc}{bc}$. This first year algebra procedure is usually taught by extending and generalizing the process of arithmetic in a formal way to operations with literal symbols. Not only should students see this parallelism of algebra and arithmetic clearly, but they should also come to appreciate the basic principles underlying the operations with natural numbers (commutative, associative, distributive laws), and

the fact that there is no reason to expect our construction of new numbers by generalization and abstraction to stop with the literal expressions of elementary algebra.

This idea of viewing one number system as an extension of another may be clarified by returning to the idea of *families* of ordered pairs mentioned briefly above. We have concerned ourselves earlier with *families* or equivalence classes of fractions. Now let us consider another set of pairs, those having 1 as the second element in each pair, for example, $(2, 1)$, $(5, 1)$, $(96, 1)$. Consider their sum, $(2, 1) + (5, 1) = [(2)(1) + (1)(5), (1)(1)] = (2 + 5, 1) = (7, 1)$. In general $(a, 1) + (b, 1) = (a + b, 1)$. In other words the sum of any two members of this family of pairs with 1 in the second position is still a member of the family. The first element of the sum of two such pairs is the sum of the two integers which were the first elements of the two original pairs. This process sets up a one-to-one correspondence between the set of all natural numbers and the set of all pairs whose second element is 1. Under this correspondence, the sum of two natural numbers corresponds to the sum of the two corresponding pairs. Mathematicians call such a correspondence an *isomorphism*. When one set of numbers is isomorphic to a part or is a subset of another set of numbers with respect to both addition and multiplication, the second number system is said to be an *extension* of the first. When we have shown that the natural numbers are isomorphic to the ordered pairs, $(a, 1)$, with respect to multiplication, we will have shown that this set of ordered pairs is a proper extension of the natural numbers, since by associating each pair $(2, 1)$, $(6, 1)$, $(a, 1)$ and so on with the natural numbers 2, 6, a, and so on we set up an isomorphism between a subset of the set of pairs and the natural numbers.

Multiplication and Division. The next step in this approach states that *by definition*, $(p, q) \times (s, t) = (ps, qt)$. If 3 is identified with the pair $(3, 1)$, this rule includes all cases such as $3 \times \frac{1}{5}$ as well as $\frac{2}{3} \times \frac{4}{5}$. The latter in this notation would be $(2, 3) \times (4, 5) = (2 \times 4, 3 \times 5) = (8, 15)$. Again, this definition was psychologically motivated by the fact that ordinary fractions had been used and operated with in this manner for centuries before modern mathematicians set out to analyze the structure of mathematics more rigorously.

We can now complete the identification with the natural numbers of the family of pairs whose second element is 1, that is, pairs of the form $(a, 1)$. We see $(a, 1) \times (b, 1) = (ab, 1)$; that is, the product of two members of this subset of all pairs is still a member of the same subset, a pair whose second element is 1. The first element of the pair, ab, is

merely the product of the first integers of each pair. This set of rational fractions is then an *extension* of the set of natural numbers under our definitions of addition and multiplication, because the natural numbers are isomorphic to a subset of the rational fractions, the set of all pairs $(a, 1)$.

Although it is not at all certain now that such an abstract approach makes a psychologically sound beginning on fractions for school children, there is some experimentation which tends to show that it may at least be acceptable at the junior high school level. This is worthy of more experimentation since the approach is in many ways more simple and more mathematically mature than are more concrete approaches.

In any event, such an abstract approach should be well within the range of superior senior high school students.

Proofs that ordered pairs obey the commutative, associative, and distributive laws when the operations are defined as above may be good exercises in algebra for superior high school students, giving them insight into the structure of the number system and into the nature of modern mathematics and its philosophy, provided that teachers point out all of these implications.

Division of fractions is probably the most difficult operation to teach. Division exists only as the inverse operation to multiplication. Questions, ideas, or problems about division must be checked by referring to multiplication. Thus,

$$(a, b) \div (c, d) = (?, ?) \text{ or } (a, b) \div (c, d) = (x, y)$$

should for thought purposes be written as $(c, d) \times (x, y) = (a, b)$. Then, by the definition of multiplication $(c, d) \times (x, y) = (cx, dy)$. If this is to be equal to (a, b), we must have $(cxb) = (dya)$. These two expressions will be equal if $x = da = ad$ and $y = cb = bc$. In other words, $(a, b) \div (c, d) = (ad, bc)$. Numerically, this says

$$\frac{3}{4} \div \frac{2}{5} = (3, 4) \div (2, 5) = (3 \times 5, 4 \times 2) = \frac{3 \times 5}{4 \times 2} = \frac{3}{4} \times \frac{5}{2}.$$

In effect, by *one* single sequence of steps this both derives and makes unnecessary the old rule that *to divide a fraction by a fraction, one inverts the divisor and multiplies.*

NUMBERS AND DIRECTION

Signed Numbers. In the elementary school, children were assured that someday they would be able *to take a larger number from a smaller number.* Students have already felt the need for some number which

would give them the answer to $5 - 8$, or a number which would describe our score in a game when we went in the hole, or a way of recording the temperature when the thermometer reading goes below zero. When we are concerned with magnitudes which can be represented by lines extending in either of two diametrically opposite directions, we have need for another kind of number, numbers which have come to be known as directed numbers, signed numbers, or positive and negative integers.

The creation of negative numbers was also motivated by the desire to make subtraction always possible and to make equations such as $x + b = a$ always solvable. Just as children comprehend that some tools are used for different purposes and that on occasion different tools may be used for the same purpose although in slightly different ways, so they learn that different numbers are used for different purposes. Historically, the use of directed numbers or signed numbers was evaded by mathematicians even after the necessity for them was known to exist. Natural numbers or integers were used when things were to be counted, fractions or rational numbers were used when things were to be measured, but negative numbers did not occur naturally in relation to concrete objects. However, one of the earliest attempts to interpret negative numbers was that of Fibonacci, an Italian of the thirteenth century. He decided that in determining profit, a negative result implied a loss. Here we have the basic idea underlying this extension of number; that is, *direction*.

Developing the Idea of Signed Numbers. As an introduction to the study of signed numbers it is useful for students to identify positive and negative numbers with such situations as an overdrawn bank account, a temperature below zero, distances above and below sea level. Their introduction to these new numbers is often not by means of their basic properties as numbers, but by means of the uses which may be made of them. The applications of positives and negatives to denote opposites are useful for motivating the study of positive and negative integers. If we were to omit all illustrations by *number lines* and problems where ideas of negative numbers may be used, negative numbers would probably appear as absurd to the modern-day high school students as they did to the early algebraists. Applications and graphical representations still play a role in gaining acceptance and understanding for new mathematical concepts.

For purposes of deriving operations and relations involving signed numbers, their principal property is not a physical idea of oppositeness but the property that every positive number, $+n$, has a negative, $-n$, such that their sum $(+n) + (-n) = 0$, and conversely, for every negative number there is a corresponding positive number with the property

that $(-n) + (+n) = 0$. Thus, 1, 5, -8, $\frac{1}{2}$, $\frac{3}{4}$ all have negatives, -1, -5, $+8$, $-\frac{1}{2}$, $-\frac{3}{4}$, such that the sum of corresponding pairs is zero.

We preserve the properties of order by ordering the integers from smaller to larger as $-n, \cdots -3, -2, -1, 0, +1, +2, +3, \cdots +n$. We will soon see that this is consistent with the earlier notion that if a positive number is added to any other number, the sum is larger than either. Here zero takes on a new role as a number. It is neither positive nor negative but the number property of an empty set, and the integer which follows -1 and precedes $+1$. Graphically it is now used to label a point or an *origin* on a line from which points corresponding to positive and negative numbers extend infinitely in both directions.

Addition and Subtraction of Integers. We can explain the addition of signed numbers in terms of counting or by addition of arrows (vectors). Thus, if the addition sign is interpreted as meaning *move to the right* and the subtraction sign is interpreted to *move to the left*, then $(+3) + (+2)$ means to start at $+3$ and move 2 units to the right. $(+3) + (-2)$ means to start at $+3$ and move 2 units to the left as in Figure 21.

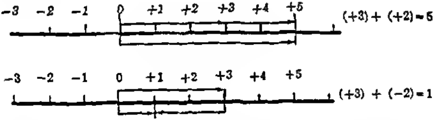


FIG. 21

A second intuitive approach to the addition of signed numbers is by an inductive process. For example, we may give students a series of problems as

$(+4) + (+3) = n \qquad (+4) + (+2) = n \qquad (+4) + (+1) = n$

$(+4) + (0) \quad = n \qquad (+4) + (-1) = n \qquad (4) + (-2) \quad = n.$

We would expect the first four problems to be done quickly and easily. The last two may cause some difficulty. Yet students can be led to see from the first four examples that as we add a smaller number to the same number the result is smaller. Hence, since -1 is one less than 0, $4 + (-1)$ should be one less than $4 + 0$, or $4 + (-1) = 3$, and so on.

A third approach, based on more sound mathematical situations,

develops the addition process out of the desire to have our associative and commutative laws apply to these numbers. For example,

$$5 + (-3) = [2 + 3] + (-3) = 2 + [(3) + (-3)] = 2 + 0 = 2$$

and,

$$-5 + 3 = [(-2) + (-3)] + 3 = (-2) + [(-3) + (+3)] = -2 + 0 = -2.$$

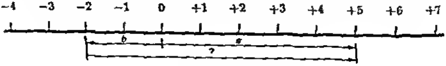
We note that in this approach we have used the fundamental idea that $(+n) + (-n) = 0$ and have assumed both the associative law and that (-5) may also be expressed as $[(-3) + (-2)]$.

Similarly in subtraction an inductive approach may be made in the following way

$$\begin{array}{ll} (+7) - (+4) = +3 & (+7) - (+1) = +6 \\ (+7) - (+3) = +4 & (+7) - (0) \quad = +7 \\ (+7) - (+2) = (+5) & (+7) - (-1) = +8 \\ & (+7) - (-2) = +9. \end{array}$$

Students observe that as the number to be subtracted decreases in size, the difference increases, and they extend this notion from $(+7) - (0) = +7$ to $(+7) - (-1) = +8$.

Let us consider the graphical approach to subtraction (Fig. 22). Remembering that subtraction is the inverse operation of addition, $(+5) - (-2)$ becomes $5 = -2 + ?$. The vector a represents 5, the sum of vector b $(b = -2)$ and vector ?. Since the sum of two vectors is the vector reaching from the *tail* of the first to the *head* of the second, the vector n which added to b will give $a = 5$ is seen to be $? = 7$.



Vector subtraction. $5 - (-2) = 7$
FIG. 22

This basic idea that subtraction is the inverse of addition may be applied directly to numerical examples as a device for leading students to discover the rules for subtraction. If a student systematically works a number of subtraction problems by converting them to addition, he will eventually (sometimes with some help) see a pattern. Of course

the examples must be planned to include the various combinations of positives and negatives such as $7 - (-2) = n$ which means $7 = (-2) + n$, $-7 - (-2) = n$ which means $-7 = (-2) + n$, $7 - (+2)$, $-7 - (+2)$, and so on.

All this leads to the generalization that to subtract a signed number we *change the sign and add.* It is interesting to note that this is basically the same as the rule that to divide by a fraction we *invert and multiply.* Both of these are special cases of a general principle that the result of an inverse operation (subtraction or division) may be obtained by changing the second number to its inverse (negative or reciprocal) and then performing the direct operation (addition or multiplication). Some people, in teaching, tacitly assume this principle and derive the *rules* from it. Logically this principle is a theorem which follows from the definitions of the inverse operations and the inverses of the elements (numbers) of the system.

The rules for *addition* and *multiplication* with positive and negative numbers are arbitrary rules, established by definition. Analogies and inductive procedures are often used to develop an understanding of the rules and of the fact that we were motivated to define them in this way to preserve the associative, distributive, and inverse element principles, but the procedures are not a substitute for an understanding of the inherent properties of numbers.

Multiplication and Division. We can interpret multiplication on the number line in a manner similar to the multiplication of natural numbers. We recall that by definition, 3×5 with natural numbers was $5 + 5 + 5 = 15$. In positive integers $(+3)(+5)$ is interpreted as $(+5) + (+5) + (+5) = +15$. A natural extension of this definition to negative integers would be $(+3)(-5) = (-5) + (-5) + (-5) = -15$. For example, if Dick loses \$5 on each of 3 days, then he has lost \$15, or $3(-5) = -15$. If we wish the commutative law to apply here as with other numbers we must define $(-5)(+3) = -15$ because $(+3)(-5) = (-15)$. By generalizing from arguments such as these we help our students to understand both that the operations with signed numbers are arbitrarily defined, and that there are reasons for choosing to define them so.

We also may use an inductive approach to motivate the definition of multiplication of signed numbers. Thus,

$$\begin{array}{r}4\\ \times 3\\ \hline 12\end{array} \quad \begin{array}{r}4\\ \times 2\\ \hline 8\end{array} \quad \begin{array}{r}4\\ \times 1\\ \hline 4\end{array} \quad \begin{array}{r}4\\ \times 0\\ \hline 0\end{array} \quad \begin{array}{r}4\\ \times -1\\ \hline -4\end{array} \quad \begin{array}{r}4\\ \times -2\\ \hline -8\end{array} \quad \text{etc.}$$

Let us look at the product of two negative numbers—how should we define $(-5)(-3)$? We know that by definition $(+3) + (-3) = 0$. By the

distributive law, if it applies, $(-5)[(+3) + (-3)] = (-5)(+3) + (-5)(-3) = -15 + (-5)(-3)$. But $(-5) \cdot 0 = 0$, hence $(-15) + (-5)(-3)$ must equal zero if signed numbers are to behave consistently with unsigned numbers. Thus we define $(-5)(-3)$ to be $+15$, since $(-15) + (+15) = 0$. More directly, in some ways, one could have written $(-5)(-3) = (-5)(4 - 7) = (-5)(4) - (-5)(7) = -20 - (-35) = +15$. This procedure assumes without proof that multiplication distributes with respect to subtraction. Through procedures such as these students come to understand *why* multiplication of two numbers of like signs is defined to give a positive result and multiplication of two numbers of unlike signs to give a negative result.

We recall that $a/b = c$ means that $bc = a$. For example, $^{15}\!/_{5} = 3$ means that $(3)(5) = 15$. Division is the inverse of multiplication. The sign rules for division are therefore rules to be derived from the multiplication rules.

$$\frac{(-15)}{(-3)} = n \text{ means that } (-3)\cdot(n) = (-15) \text{ and that } n = +5.$$

$$\frac{(+15)}{(-3)} = n \text{ means that } (-3)\cdot(n) = (+15) \text{ and that } n = -5.$$

$$\frac{(-15)}{(+3)} = n \text{ means that } (+3)\cdot(n) = (-15) \text{ and that } n = -5.$$

$$\frac{(+15)}{(+3)} = n \text{ means that } (+3)\cdot(n) = (+15) \text{ and that } n = +5.$$

Here again concrete, physical analogies, diagrams, and induction have been used in the early stages of development, but basic principles are not ignored but pointed out repeatedly and concretely.

OPERATIONS WITH SIGNED NUMBERS AS ORDERED PAIRS OF RATIONAL NUMBERS

Let us look at the ideas of number represented by the symbols, 5, +5, and −5. Although we have associated them with two (or three) *different* kinds of number, they have many common properties. We recall that $(3 + 2)$, $(4 + 1)$, $(1 + 4)$, $(6 - 1)$, $(7 - 2)$ all belong to the class 5, a natural number. We recall that $^2\!/_4$, $^4\!/_8$, $^6\!/_{12}$, $^8\!/_{16}$ $\cdots$ all belong to the equivalence class, $\frac{1}{2}$, a fraction.

From this viewpoint we can regard +5 as the symbol for the set of ordered pairs of numbers such that the first minus the second is 5, such as (8, 3), (6, 1), (7, 2), (10, 5), (115, 110). Extending this idea further

(3, 8), (1, 6), (2, 7), (5, 10), (110, 115) $\cdots$ all belong to the class of ordered pairs of numbers such that the second minus the first is 5. We can associate this class with -5, a negative number.

At advanced levels of thought two ordered pairs of rational numbers (a, b) *and* (c, d) *are defined to be equivalent if and only if* $a + d = b + c$. Of course, the thinking that led to this definition was based on the fact that with natural numbers if $a - b = c - d$, then $a + d = b + c$. For example, $(7, 2) = (9, 4)$ because $7 + 4 = 11 = 2 + 9$. In this case we could also have written $7 - 2 = 9 - 4$. However, with natural numbers we could not have assigned any meaning to $2 - 7$ or $9 - 4$, but with ordered pairs we may also write (2, 7) and (4, 9) and show that they too are equal because $2 + 9 = 7 + 4$.

We can now compare the ordered pair approach to signed numbers with our earlier definitional approach. By our definition of equality all pairs with both members the same are equal, that is $(a, a) = (b, b)$ because $a + b = a + b$. Further, any pair of this type corresponds to zero because when added to any other pair the new pair is equal to the original. To test this we must first define addition of pairs. By definition $(a, b) + (c, d) = (a + c, b + d)$ or $(2, 5) + (9, 2) = (11, 7)$. This was motivated by and corresponds to the fact that $(2 - 5) + (9 - 2) = 11 - 7$. In ordered pair notation $(5 + k, 0 + k)$ corresponds to $(+5)$ and $(0 + m, 5 + m)$ corresponds to (-5). Then $(+5) + (-5)$ corresponds to $(5 + k, 0 + k) + (0 + m, 5 + m) = (5 + k + 0 + m, 0 + k + 5 + m)$. Since the two elements of this last pair are identical, the pair corresponds to zero and the sum of the ordered pairs corresponds to $(+5) + (-5) = 0$. We have essentially proved that for addition the set of positive and negative integers, and zero, $\cdots -3, -2, -1, 0, +1, +2, +3, \cdots$ is isomorphic to the set of classes of equal ordered pairs of integers under these definitions of addition and equality. Note that corresponding to every integer as ordinarily written there is an infinite set of pairs, thus -5 corresponds to $(0 + m, 5 + m)$ for all values of m. Any member of this set could be used to represent it and would behave the same as any other of this set. This compares with our previous work with *fractions* where $\frac{1}{2}$ or (1, 2) represented an infinite family $(1m)/2m$ or $(1m, 2m)$ all of which were equal and behaved alike.

After we define multiplication of these new pairs corresponding to signed numbers we could prove them isomorphic to the usual set of integers with respect to multiplication too. We could then go one step further and show that the set of all pairs $(a, 0)$ whose second element is zero is isomorphic under both addition and multiplication to the set

of all natural numbers and zero. This would show that this system of pairs is an extension of the system of integers. Further if we had used *rational number* (and *irrational number*) wherever we used *integer* above, the arguments would have been the same; that is, the (positive) rational and irrational numbers can be extended in this same manner to a system of signed rational and irrational numbers, or, in other words, to the set of all real numbers.

At this level of abstraction we define the product of two signed numbers when they are represented by ordered pairs, as

$$(a, b)(c, d) = (ac + bd, bc + ad).$$

By way of illustration with the same examples as we used when they were represented by single numbers with $+$, $-$ signs this definition would give:

$$(+5)(+3) = (5, 0)(3, 0) = (15 + 0, 0 + 0) = (15, 0) = +15$$

$$(-5)(+3) = (0, 5)(3, 0) = (0 + 0, 15 + 0) = (0, 15) = -15$$

$$(+5)(-3) = (5, 0)(0, 3) = (0 + 0, 0 + 15) = (0, 15) = -15$$

$$(-5)(-3) = (0, 5)(0, 3) = (0 + 15, 0 + 0) = (15, 0) = +15.$$

We define division by the equation

$$(a, b) \div (c, d) = (x, y) \quad \text{if and only if} \quad (x, y)(c, d) = (a, b).$$

Following the patterns of our previous discussion with rational numbers we would derive a rather elaborate expression for (x, y) which, since it is less intuitively obvious and less useful than other forms, we will leave to the reader to work out (f)

Since the rules for operating with negative numbers are applicable to all rational numbers, not merely to the integers which we have used in our illustrations, the operations of subtraction and division are now always possible in the field of rational numbers (with the exception of division by zero) and all linear equations are now solvable. The categories of numbers still needed in elementary mathematics are the irrationals and the complex numbers.

BEYOND THE RATIONALS

Irrational Numbers. Using only the rational numbers we are unable to solve some quadratic equations such as $x^2 = 2$ which may arise from such a simple geometric problem as the determination of the length of the hypotenuse of a right triangle.

This incompleteness in our number system may be pointed out to students as early as when repeating decimals are first met. We noted earlier that each rational number is represented by either a terminating or a repeating infinite decimal. There also are infinite decimals which neither terminate nor recur. For example we can construct the number .101001000100001... where each '1' is followed by a set of zeros containing one more zero than in the set preceding the '1'. Such an infinite decimal would never repeat and never terminate.

A more familiar though less obvious example of a nonrepeating, nonterminating decimal is found in $\pi = 3.14159\ldots$. By this we mean that there is a succession of rational numbers; namely, $r_1 = 3$, $r_2 = 3.1$, $r_3 = 3.14$, $r_4 = 3.142$, $r_5 = 3.1416$, $\cdots$ which differ from π by smaller and smaller amounts such that r_n can be made to be as good an approximation to π as is desired by taking n sufficiently large.

The ancient Greeks realized and proved that there was no rational number whose square is 2. But such numbers as the square root of 2 are required for geometry. If a square is 1 unit on each side, then the length of the diagonal of that square is a number whose square is 2. It was customary with the Greeks, as it is today, to express lengths or distances in terms of whole numbers and the ratios of whole numbers, rational fractions, but whole numbers and their ratios were closely tied in with the philosophy of the Pythagoreans. Consequently they were disturbed to discover the existence of a constructable but "incommensurable" distance, the diagonal of a square (Fig. 23). Just as we have con-

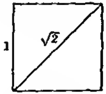


FIG. 23

structed numbers to solve such equations as $4 - 7 = n$ or $3n = 10$, so additional numbers had to be constructed to be solutions of conditions such as $x^2 = 2$ and to represent lengths such as diagonals of squares. Today there are several satisfactory ways of doing this, but all of them have been developed within the last century. Symbolically we represent the solution of $x^2 = 2$ by $\sqrt{2}$, or in general, a solution of $x^n = c$ by $\sqrt[n]{c}$. Formally, d is a member of the solution set of the condition $x^a = b$ if and only if $d^a = b$. The numbers represented by these symbols must be constructed from the set of rational numbers just as our rational numbers were constructed from the integers. This construction is different from those previously encountered and involves the concept of limit.

Actually children meet the idea of a limit early and fairly often. When we write $\frac{1}{3} = .3333\ldots$ we mean that $\frac{1}{3}$ is the limit of the sequence of numbers .3, .33, .333, .3333, and so on, which may be derived by taking the partial sums of the series $.3 + .03 + .003 + .0003 + \cdots$. Children may be fascinated by recurring designs such as in Figure 24 in which the area of the innermost square quite clearly approaches zero, and the sum of the areas of the squares considered separately, if the larger square is one unit on a side, is $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \cdots$. This infinite sum can be written as the sequence $1, 1\frac{1}{2}, 1\frac{3}{4}, 1\frac{7}{8}, 1\frac{15}{16}, 1\frac{31}{32}, \cdots$, where the second, third, and nth terms of the sequence represent, respectively, the sum of 2, of 3, and of n terms of the series, a geometric progression normally studied in detail in second year algebra. However, this sequence, and hence the series, can be seen intuitively by inspection or by drawing diagrams of the successive partial sums to approach 2 as a limit without having recourse to the formal processes for summing a geometric progression.

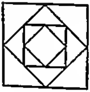


FIG. 24

Although, as stated, children meet the idea of a limit early and are fascinated by it, both the concept and its rigorous treatment are moderately sophisticated. Comprehension of such concepts develops out of repeated contacts with them at successively more mature, general, and abstract levels. For these reasons we feel that teachers should seek opportunities to introduce these ideas wherever appropriate but should not devote too long a period to them nor expect complete insight and mastery in the students' early contacts with them. For example, in recent years it has often been argued that such ideas, though logically unavoidable, are too mature to be introduced into the discussion of area even in demonstrative geometry. As a result of this some teachers have, when deriving the formula for the area of a rectangle, avoided discussions of the possibility that the sides might be incommensurable with each other or with a unit of measure. With the restrictions noted above and for the reasons stated there, we believe that the limit concept should be introduced here and wherever appropriate. However, we, in this chapter, must turn back from the general concepts of limit to its relationships to *number*, especially to ***irrational numbers***. For a detailed discussion of the limit concept see Chapter VII of the twenty-third yearbook.

An irrational number, a number which may not be written as an ordered pair of integers, may be defined as the limit of an infinite sequence of rational numbers in many ways. Perhaps the most obvious example and most direct process occurs when we slightly formalize the *division* method for finding the square root of two. Testing quickly shows that since $1^2 = 1$ and $2^2 = 4$, the square root of two is between the integers 1 and 2. This assumes that certain properties of order exist in the rational number system, for example, that if a, b, c are positive and $a > b$ then $ac > bc$ and $a^2 > b^2$. We will not take the time to elaborate these principles of order and the associated inequalities here, but we strongly recommend them to our readers for further study and for incorporation into the teaching of algebra, geometry, and analysis in the high school and perhaps in some junior high classes.

This first fact can then be written as $1 < \sqrt{2} < 2$ and the interval from 1 to 2 can next be subdivided into n parts. In our decimal number system it would be convenient to take $n = 10$. By testing 1.1, 1.2, ... 1.8, 1.9 we find that $1.4 < \sqrt{2} < 1.5$. Repeating this subdivision and testing process will show successively that $1.41 < \sqrt{2} < 1.42$, $1.414 < \sqrt{2} < 1.415$, and so on. Each of these intervals 1 to 2, 1.4 to 1.5, 1.41 to 1.42, 1.414 to 1.415 is *nested* within all the preceding intervals and the set of nested intervals gives rise to two sequences: the sequence of lower bounds, 1, 1.4, 1.41, 1.414, $\cdots$, and the sequence of upper bounds 2, 1.5, 1.42, 1.415, $\cdots$. If it can be shown (as it can be in this case) that these two sequences approach the same limit, that limit is by definition the square root of two. Equivalent to the proof that the two sequences have the same limit would be a proof that the length of the intervals in our nested sequence approaches zero as a limit. Intuitively the latter approach is helpful because one can conceive of a point (Fig. 25) which is within all of the intervals. This would be the point corresponding to $\sqrt{2}$. It could be constructed with a straight edge and compasses (see Figure 23) but could never be located by measuring with a ruler.

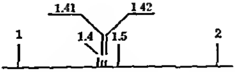


FIG. 25

Every irrational number can be described as the common part or intersection of a set of nested intervals whose length approaches zero as a limit. Given two sets of nested intervals, the sum of the corresponding numbers can be defined to be the number represented by a new

set of nested intervals derived from the two given ones. If the number M is defined by the intervals $a_1 \leq M \leq b_1$, $a_2 \leq M \leq b_2$, $a_3 \leq M \leq b_3$, $\cdots a_i \leq M \leq b_i$ and a second number N is defined by the set of nested intervals $c_i \leq N \leq d_i$, then $M + N$ can be defined by

$$(a_1 + c_1) \leq (M + N) \leq (b_1 + d_1), (a_2 + c_2) \leq (M + N) \leq (b_2 + d_2), \cdots (a_i + c_i) \leq (M + N) \leq (b_i + d_i), \cdots.$$

Multiplication of numbers defined by nested intervals can be similarly defined and on these definitions of addition and multiplication we can base definitions of their inverses, subtraction and division.

This system for defining numbers actually describes the rational numbers, too, and hence can be thought of as including the entire system of real numbers (irrationals, rationals, integers, natural numbers). This is easily demonstrated by remembering that $\frac{1}{3} = .33\bar{3}\ldots$ can be regarded as $.3 \leq \frac{1}{3} \leq .4$, $.33 \leq \frac{1}{3} \leq .34$, $.333 \leq \frac{1}{3} \leq .334$, $\cdots$. Similarly, although numbers such as .5 are called terminating decimals, they too can be written as infinite repeating decimals which can then be translated into nested intervals. Thus, .5 can be written as a repeating decimal in two ways, $.5 = .500\bar{0}\ldots = .499\bar{9}\ldots$. The second of these says $.4 \leq .5 \leq .5$, $.49 \leq .5 \leq .50$, $.499 \leq .5 \leq .500$, $\cdots$ and so on. All of this is really unnecessary in a sense, because we know that the upper bound, .5, is the number we are describing exactly, but it does show that nested intervals can be used to define both rational and irrational numbers. The property of being represented by *repeating sequences* of digits in a place value numeration system with *any base*, $b > 1$, is a property which distinguishes rational numbers from irrationals independently of the base. Both rationals and irrationals, and hence all real numbers, can be represented by the nested interval technique.

There are two other currently used procedures for defining irrational numbers out of sets of rationals. George Cantor defined real numbers directly as the limits of sequences of rational numbers. A sequence of rational numbers $a_1, a_2, a_3, a_4, \cdots a_n, \cdots$ is defined to have a limit, A, if for every positive number ϵ there exists an N_ϵ such that for all terms in the sequence after the N_ϵth the difference between A and a_n is less than ϵ, that is for $n > N_\epsilon$, $|A - a_n| < \epsilon$. Using this definition it can be proved as a theorem that a necessary and sufficient condition for such a limit to exist is that the difference between two terms of the sequence approaches zero as a limit, as the two terms are selected farther and farther along in the sequence. Such sequences are sometimes called *Cauchy sequences*. A more technical statement of the theorem would

then he: *A sequence of rational numbers, $a_1, a_2, a_3, a_4, \cdots a_n, \cdots$ is a Cauchy sequence and represents a real number if for every $\epsilon > 0$ there exists an N_ϵ such that for all n, m greater than N_ϵ, $| a_n - a_m | < \epsilon$.* Since every infinite decimal fraction can be regarded as a sequence of rational numbers which can be shown to approach a limit, every infinite decimal represents a real number. From this viewpoint $\frac{1}{3}$ can be represented by .3, .33, .333, $\cdots$ and .5 can be represented by either .4, .49, .499, $\cdots$ or by .5, .5, .5, $\cdots$. Thus this Cantor sequence approach also includes both rational and irrational numbers in its definition of the real numbers.

The third approach to irrationals by way of *Dedekind cuts* is closely related to the nested interval approach but makes more use of set theoretic ideas. We shall not take the space to expound it or other approaches such as that via continued fractions.

Transcendental Numbers. One of the first irrationals to be met in this chapter, and in the historical development of mathematics as well as in the secondary school curriculum, is the number today denoted by the Greek letter π. This number is a member of a subset of the irrationals which is more numerous (in the sense of infinite sets) than the set of all nth roots of rationals, but which was not even proven to exist until 1844. This is the set of numbers called *transcendental*. This name is used for all numbers which are *not* algebraic where *algebraic numbers* are defined to be numbers which can be roots of equations of the form $a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0$ where n is a positive integer and all the a_i are integers. Thus $\sqrt{2}$ and $\sqrt[3]{2}$ as well as $-\frac{2}{3}$ are algebraic because they are the roots respectively of $x^2 - 2 = 0$, $x^3 - 2 = 0$, and $3x + 2 = 0$. Other transcendental numbers met in elementary mathematics in addition to π are e, the logarithms of many numbers, and the trigonometric functions of many angles. The fact that π was not proven to be transcendental (and the famous Greek problem of squaring the circle finally thereby shown to be impossible of solution with straight edge and compasses) until 1882 shows something of the difficulties of a theoretical treatment of irrationals and transcendentals. This, however, does not mean that secondary school students should not have at least heard of them. Quite the reverse! Here is one of the few places where secondary school students can be brought close to the boundaries of modern mathematics and helped to feel something of its vigorous vitality.

Computation with Irrationals. In practical computations irrational numbers are replaced by rational approximations such as $\pi \approx 3\frac{1}{7}$, or $\pi \approx 3.14159$, or $\sqrt{2} \approx 1.414$. The topic of computation with numbers arising from approximations is discussed later in Chapter 5.

The Real Numbers. In conclusion we will merely point out again that although we have used the decimal place value notation system, especially decimal fractions, in discussing irrationals, the basic distinctions between rationals and irrationals are not properties of any particular base nor even of the general idea of place value number notation. However, the limit concept in some form is inescapable in defining irrationals and can be applied in such a way that rationals and irrationals all are seen to be part of the same real number system.

THE LAST (?) NEW NUMBER

With the real numbers we can solve all equations of the form $ax + b = 0$ and of the form $cx^n + d = 0$ where n, b, c, d are real numbers and where c and d have opposite signs. However, such a simple little equation as $x^2 + 1 = 0$ can not be solved using any of our real numbers. There are many more equations of interest to both pure and applied mathematicians which can not be solved using only the real numbers. To end this unhappy situation another new number was defined. In elementary approaches this is commonly represented by the letter i, and i is endowed by definition with the property that $i^2 = -1$. This definition suffices to show that i is a solution of $x^2 + 1 = 0$. We will assume that our readers are familiar with the further definitions whereby this *unit imaginary*, i, is combined with real numbers to produce *complex numbers* such as $2 + i$, $3 - 2i$, $\sqrt{2} + 1.5i$, and so on. The general complex number is defined to be any number of the form $a + bi$ where a and b represent real numbers.

The equality, sum, difference, product, quotient, and integral powers of complex numbers, $a + bi$, are then usually defined with such a strong appeal to their analogy with real binomials as to make these *definitions* appear to students to be *derivations*. Thus $(a + bi) \cdot (c + di)$ is often *multiplied out* to give $ac + adi + bci + bdi^2$, then i^2 is replaced by -1 and the final result written as $(ac - bd) + (ad + bc)i$. Actually, this latter expression is a *definition* of $(a + bi) \cdot (c + di)$, but many students never even see it. They are taught to use the analogy with multiplication of binomials in every numerical situation such as $(2 + 3i)(-4 - i)$ and the general definition is pointed out passingly if at all.

Pedagogically this approach has the values of being concrete, capitalizing on previous experiences, using a (partially) meaningful process rather than an arbitrarily presented and memorized rule. This presentation fails to explain the logical foundations of these numbers and their properties. It fails to show clearly their fundamental differences from the real numbers as well as their similarities and the sources of these similarities. In the writers' opinions, an approach which omits such

fundamental ideas or leaves them unnecessarily obscured is not a fully sound pedagogical approach. We shall here present concisely a modern abstract approach to complex numbers and then suggest considerations from which our readers may develop a completely sound pedagogical approach. We shall be quite concise because the details are so much like those we have elaborated in connection with the integers and the rational numbers.

In this approach a complex number is defined to be any ordered pair of real numbers. Thus if a and b are real numbers, (a, b) is a complex number. We then give the following arbitrary definitions:

Equality: $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$

Addition: $(a, b) + (c, d) = (a + c, b + d)$

Multiplication: $(a, b) \times (c, d) = (ac - bd, ad + bc)$.

Note that each of these definitions for ordered pairs is phrased in terms of previously established numbers (real numbers) and operations with them. Using the idea that we still want to define subtraction and division to be the inverses of addition and multiplication we can work out the formulas:

$$(a, b) - (c, d) = (a - c, b - d)$$

$$(a, b) \div (c, d) = \left(\frac{ac + bd}{c^2 + d^2}, \frac{bc - ad}{c^2 + d^2}\right).$$

These formulas are actually based on the idea that the pair (complex number) $(0, 0)$ when added to (a, b) gives (a, b) and that the pair $(1, 0)$ when multiplied times (a, b) gives (a, b). These pairs, $(0, 0)$ and $(1, 0)$, are called the *additive* and *multiplicative identities* respectively. These pairs correspond in our *new* (complex) number system to the *zero* and *one* of our integers.

We could now prove that the complex number system also obeys the *closure, associative, commutative,* and *distributive* laws which we have preserved at each step of the process of building new numbers. In fact, although logically we could have defined equality, addition, and multiplication of complex numbers any way that we wished, we were actually guided in our choice of definitions by the desire to have our new numbers again behave as much like our old ones as is possible. Actually the complex numbers lack one property which is a most important and useful one in the real field. This is the property of *order*. For any two real numbers a and b it is always true that $a > b$, or $a = b$, or $a < b$. The ideas of *greater than* and *less than* can not be defined for complex numbers in such a way as to have the same properties as these do in the real

numb... up something when we extended the real numbers to the complex field. In this sense then the real numbers were the *last numbers*—i.e., the last numbers with all those properties.

However, we also gained something; namely, in the field of complex numbers all equations of the form $a_0x^n + a_1x^{n-1} \cdots a_{n-1}x + a_n = 0$, where the a_i are real numbers, not only have a solution (as proven by C. F. Gauss about 1800), but in fact have n roots. Actually this same statement would be true even if the coefficients, a_i, were complex. It is for this reason that the complex number system is called *complete*—i.e., no new numbers are needed to solve the general equation even if the coefficients are themselves complex.

Since we are assuming that those who have read this far already know about the traditional graphical representation and polar forms of complex numbers and their fascinating relationships with the trigonometric functions (if you do not, see the references in Chapter 11), we will not expound on them or on their applications. (Remember, in the study of electricity our unit imaginary is represented by j rather than i.)

However, pedagogically as well as historically, these complex numbers are probably the hardest ones for students to accept. The name *imaginary* is both merely a historical hangover and a pedagogical handicap as it seems to indicate that there is something almost supernatural about them. It would be better to stress the term *complex*, but one cannot omit mentioning *imaginary* since it is so generally used. Historically and pedagogically, making graphical representations of both the numbers and the operations with them and some discussion of their many and varied applications will quicken their acceptance. All of these together with some later discussion of the logical structure of number systems and the reasons for and methods of extending them will help students ultimately to develop in their understanding not only of these numbers but of number systems and mathematics as a whole.

We cannot leave the topics of numbers and operations without a word or two on each of three topics—(1) still further new numbers, (2) the role of structure in mathematics, and (3) the connections between operations and relations.

STILL OTHER NEW NUMBERS

We can make no further extensions of our number system without making some sacrifices. In fact, as we noted above, we had to sacrifice the order relations *greater than* and *less than* when we defined the complex numbers. However, if we don't mind an occasional sacrifice there are many sets of mathematical elements for which we can define operations

comparable to addition and multiplication. Some examples are: *quaternions* which can be represented by quadruples of real numbers rather than by mere pairs, *vectors* which can be regarded as special cases of matrices and which can be defined as triples of real numbers, *matrices* which are rectangular arrays of real numbers, *Gaussian integers* which are numbers of the form $a + bi$ where a and b are restricted to being ordinary integers, *algebraic integers, p-adic numbers,* and others. Quaternion, vector, and matrix algebras all sacrifice the commutative law for multiplication, i.e., it is not necessarily true in these algebras that $A \times B = B \times A$. Gaussian integers are much like our ordinary integers (which are also called *rational integers*). Thus, though 5, which is prime in the set of ordinary integers, is $(2 + i) \times (2 - i)$ in Gaussian integers, factorization is unique in Gaussian integers. However, if we call *integers* the set of all numbers of the form $a + b\sqrt{-5}$ where a and b are ordinary or *rational* integers, we find that these new integers fail to satisfy the fundamental theorem of arithmetic; namely, that every integer can be uniquely (except for sign and order) factored into prime factors. In this last system, for example, 21 can be factored into primes in two different ways:

$$21 = 21 + 0 \cdot \sqrt{-5} = 3 \cdot 7 = (1 + 2\sqrt{-5}) \cdot (1 - 2\sqrt{-5}).$$

These arithmetics and algebras, and many more, have been and will be further explored by many mathematicians. There is no end to mathematics, its fascination, and its uses!

STRUCTURE IN ARITHMETIC AND ALGEBRA

Two simple finite arithmetics are defined by the two sets of tables in Figure 26. These tables mean, for example, that in Arithmetic I, $b + c = d$, and in Arithmetic II, $b \times c = c$.

It is an interesting exercise for oneself or for a bright group of students to look for, tabulate, analyze, and systematize the similarities and the differences between Arithmetic I and Arithmetic II, between addition and multiplication in general, between addition in the two different arith-

Arithmetic I

+	a	b	c	d	e
a	a	b	c	d	e
b	b	c	d	e	a
c	c	d	e	a	b
d	d	e	a	b	c
e	e	a	b	c	d

×	a	b	c	d	e
a	a	a	a	a	a
b	a	b	c	d	e
c	a	c	e	b	d
d	a	d	b	e	c
e	a	e	d	c	b

Arithmetic II

+	a	b	c	d	e	f
a	a	b	c	d	e	f
b	b	c	d	e	f	a
c	c	d	e	f	a	b
d	d	e	f	a	b	c
e	e	f	a	b	c	d
f	f	a	b	c	d	e

×	a	b	c	d	e	f
a	a	a	a	a	a	a
b	a	b	c	d	e	f
c	a	c	e	a	c	e
d	a	d	a	d	a	d
e	a	e	c	a	e	c
f	a	f	e	d	c	b

FIG. 26

metics, multiplication in the two. For example, (1) all four tables have the property that none of the entries within the table uses a number or symbol which did not occur in the outer row and column. This property is called *closure*, which means that the sum (product) of two elements in the basic set is also an element in the set. Other things which may be observed are that: (2) whatever two elements you pick, x and y, $x + y = y + x$ and $x \times y = y \times x$, (3) in each table there is a row (and column) which is exactly the same as the top or outside row (left-hand column). Property (2) is called the *commutative law*, and property (3) is expressed by saying that there is an *identity element* for addition (a in these examples) and one for multiplication (b in these examples). Thus if x is any other element in the above sets $x + a = x$ and $x \cdot b = x$. There are other similarities and differences. For example, subtraction can be defined for all pairs in both arithmetics, but division can be completely defined only in Arithmetic I. Can you define these operations and explain why and in which cases one cannot completely define division in Arithmetic II? Actually, these are examples of *modular arithmetics*, modulo 5 and 6, and can be written with numbers, representing a, b, c, d, e, f by 0, 1, 2, 3, 4, 5. Try it! Can you figure out an explanation for why division will be always possible only in arithmetics with a prime modulus?

These arithmetics, though fascinating, are not to be explored here. They are exhibited to illustrate the point that there are certain basic ideas and relationships that occur time and again in mathematics, in many different situations, sometimes disguised by different symbols and terms, but still fundamentally the same ideas. If these basic ideas can be perceived to be present in a new mathematical structure, the mathematician can then immediately write down for the new system the consequences which he has previously shown must always follow when the basic structure, however disguised, is present. For the student the perception of this structure increases his understanding of the new structure and also of why it was originally defined and built as it was.

There are many different typical mathematical structures. Two very important ones which are also present in our arithmetic and algebra are those called *groups* and *fields*.

A *group* is any system having a set of elements with a *single* operation and equality defined such that it

(1) *is closed*.
(2) *is associative*.
(3) has an *identity element*. (In ordinary arithmetic this is *zero* for addition and *one* for multiplication. It is a for addition and b for multiplication in Arithmetics I and II.)

(4) for every element, has an *inverse element*. (An element such that added to (or multiplied by) the given element their sum (or product) is the identity element. Thus the additive inverse of $+2$ is -2 and of -3 is $+3$ in ordinary arithmetic, while the multiplicative inverse of 2 is $\frac{1}{2}$ and of $\frac{3}{4}$ is $\frac{4}{3}$. In Arithmetic I the additive inverse of d is c, in Arithmetic II it is d itself.)

If the elements of the group are also commutative the group is called a *commutative* or *Abelian* group. Thus the positive, negative, and zero integers form an Abelian group under addition. The positive rational numbers form a group under multiplication. Arithmetics I and II each form groups under addition, but II is not a group with respect to multiplication. Can you tell why?

A *field* is a system having a set of elements with *equality* and *two* operations defined such that:

(1) It is a commutative group with respect to addition.
(2) It is a commutative group with respect to multiplication (if the additive identity, zero, is omitted).
(3) Multiplication is distributive with respect to addition, that is $a \cdot (b + c) = ab + ac$.

Thus the rational numbers, the real numbers, and the complex numbers are fields. So is Arithmetic I. We could have said that in this chapter we were going to study some of the groups and fields of modern mathematics. We felt that here, as in secondary classrooms, abstractions and generalizations should, in general, be built up from beginnings which are more concrete and special. However, we have tried to show the importance of these ideas as we went along and how they are a continuous thread running though all our arithmetic and algebra. In this section we have merely hinted at the fact that they are a part of the structure of many other mathematical systems and that as such they are most useful and fruitful concepts for the advanced mathematician and for the students whose insights and understandings are increased and broadened as he perceives them.

OPERATIONS AS RELATIONS

In the following chapter, "Relations and Functions," a relation will be defined as a set of ordered pairs of elements. The meaning, significance, use, and pedagogical advantages of the concept of relation will be discussed and illustrated there. Hence we will merely note that operations may be classified according to the number of elements involved, as unitary, binary, ternary, and so on, and further, they may all be classified under the heading of relations. Thus ordinary addition of integers may

be regarded as a binary operation and as a set of ordered pairs in which the first element of the pair is itself a pair of integers and the second element of the pair of the relation is a single element, their sum. This can be represented as illustrated in Table 4. The meaning of the *domain* and *range* will also be further illustrated in the next chapter.

TABLE 4
RELATION "+" DEFINED FOR INTEGERS

First element	Second element
Domain: pairs of integers	Range: integers
(2, 3)	5
(−3, 4)	1
(0, 2)	2
(−4, −5)	−9
(−5, −4)	−9
(0, 0)	0
etc.	etc.

In set notation, this relation can be represented by

$$\{[(2, 3), 5], [(-3, 4), 1], [(0, 2), 2], [(-4, -5), -9],$$
$$[(-5, -4), -9], [(0, 0), 0] \cdots\} \text{ or by } \{[(x, y), z] \mid x + y = z\}.$$

This approach to the concept of an operation is interesting and enlarges one's mathematical horizons, but is beyond the scope of this chapter and probably not, relatively, a significant concept for the secondary school. *See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in the chapter.*

NOTES

1. Bell, E. T. *The Development of Mathematics.* New York: McGraw-Hill Book Co., 1940. p. 3.
2. Fraenkel, Abraham A. "Integers and Theory of Numbers." *Scripta Mathematica* 12–17; 1955.
3. *Illustrated London News.* October 2, 1957. p. 550–53.
4. Brownell, William A. *Arithmetic in Grades I and II.* Durham, N.C.: Duke University Press, 1941. p. 11–63
5. Buckingham, Burdette R., and MacLatchy, Josephine H. "Number Abilities of Children When They Enter Grade One." *Report of the Society's Committee on Arithmetic.* Twenty-Ninth Yearbook, National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Company, 1930. p. 473–524.
6. Zwick, J. S. "What Does the Kindergartner Know About Numbers." *Ohio School* 28: 263; September 1950.
7. Corwin, Bjonerud. *Arithmetic Knowledges of Kindergarten Children at Time of Admission to School.* Unpublished dissertation. Detroit: Wayne State University.

3

Relations and Functions

KENNETH O. MAY AND HENRY VAN ENGEN

"The cultivation of the mathematical imagination depends chiefly on the child being put into the right attitude towards mathematical conceptions in his earliest years; and, after that, on the right use being made of certain nodes or critical points, which occur here and there in each branch of mathematics and which should be dealt with in a quite different manner from the rest of the course."—MARY EVEREST BOOLE in *Preparation of the Child for Science.* 1904.

NUMBER PAIRS IN ELEMENTARY AND SECONDARY MATHEMATICS

TEACHERS of mathematics are very familiar with the words 'relation' and 'function'. Both words are frequently used in a nonmathematical context as well as in a mathematical context. In most instances, 'relation' is used very loosely and with many shades of meaning. In fact, in elementary and secondary mathematics 'relation' has had no precise meaning.

The word 'function' has had a precise meaning in secondary mathematics for some time. However, it is frequently poorly understood and, even more frequently, not even mentioned as an important concept in ninth grade algebra.

Recently mathematical thinking has brought about a change in the way these words are used. 'Relation' is now given a very definite mathematical meaning, and there has been a change in the definition of 'function'—a very fundamental change. Contemporary mathematics prefers to define 'relation' and 'function' in terms of sets; in fact, in contemporary mathematics the idea of set is considered to be more fundamental than the idea of number.

Sets of numbers and sets of number pairs appear very often in daily affairs. Familiar examples are: height-weight charts, temperature-time charts, and cost-of-living charts. It is easy to think of other cases in which we deal with pairs of numbers. For example, we can pair the weight (number of ounces) of a letter with the postage (number of

cents) it requires. We often pair objects other than numbers. In the grade book, a final average is paired with each student at the end of the year. Also, in a class record book we pair with each hour the names of students in the class at that hour.

Perhaps the first pairing that the small child encounters is the *association of names with objects*. He soon finds that more than one object is paired with the same name, and that more than one name is often associated with the same object. In his first arithmetical experiences the child learns to count a collection of objects by pairing off the objects with names, 'one', 'two', 'three', and so on. He learns that by this counting process a number name is associated with each collection; that is, one can pair off collections with number names. In these and many other instances the child learns to pair things, that is, to associate things, to make things correspond. Of course, he does this without being consciously aware of the idea of pairing.

In the schools, we first study number pairs in tabular and graphical form. The elementary school pupil encounters tables and graphs as ways of showing how numbers are paired. This first encounter may be a time-temperature graph in which he has recorded the temperature for each hour of the school day. In this case the graph visually shows such number pairs as: (8, 45°), (9, 47°), (11, 55°), (12, 56°), (1, 60°), (2, 61°), (3, 59°), (4, 54°). Here the first member of each pair records the hour of the day and the second element records the temperature. You are familiar with graphs, and hence, we will not show the pairings as they might be presented in elementary school or junior high school.

The elementary school pupil also meets pairings involving things other than numbers. For example, (Iowa, 2,600,000), (Kansas, 2,000,000), (Illinois, 9,400,000), (Minnesota, 3,200,000), (Nebraska, 1,400,000). Here states are paired with their populations. These are the raw data from which the pupil could make a graph. It is a set of pairs which lends itself to tabulation and graphical techniques.

Since these pairing situations occur so frequently, it is not surprising that the study of pairs is of great importance in mathematics. We shall see that the pairing idea is the key to understanding many important mathematical ideas.

Because tables and graphs are simple ways of visualizing pairings, they are suitable for use in the elementary grades. After working with pairings as shown in tables and graphs, the pupil is ready for a more abstract method of showing how numbers are paired. The pupil learns that it is easy to write '$A = s^2$' as a rule for pairing certain numbers. The formula, or rule, is more concise than the table of squares appear-

ing in every student's mathematics book which expresses the same pairing in tabular form.

In many other situations the pupil learns to use formulas along with tables and graphs, and much time is spent in achieving an understanding of these different devices for showing how numbers are paired. As soon as the pupil has reached this stage, he is already dealing with some features of mathematical relations; namely, (1) a collection of pairs (which may be exhibited in a table or graph), and (2) a rule or formula that indicates how the objects are paired. In our example above we have a collection of pairs, $\{(1, 1), (2, 4), (3, 9), \cdots\}$ and a rule $A = s^2$, which tells us that the second number in each pair is the square of the first.

Too frequently, in algebra classes, the rule is stressed almost to the exclusion of such other aspects of relations as number pairs and sets of number pairs. But there is need to keep all of these before the student at all times. This is certainly desirable because both the formula and the number pair idea are important. The most effective work in mathematics comes from using them simultaneously. However, in spite of the usefulness of formulas, it is the pairs of objects with which we are basically concerned. The formula is just a way of defining, describing, and dealing with a collection of pairs. Many different formulas give the same collection of pairs. For example, $A = s^2$, $A - s^2 = 0$, and $s = \pm\sqrt{A}$ all define the same collection of pairs $\{(1, 1), (-1, 1), (2, 4), (3, 9), \cdots\}$. Since we are concerned in this chapter with the nature of relations and the means by which the student may be brought to understand them, we shall approach the subject by considering collections, or sets of pairs of numbers.

SETS AND VARIABLES

The idea of a set of things is very common and quite easy to grasp. Children think in terms of sets and operations with sets at a very early age. Mathematicians have taken the notion of set as a very basic element in the construction of virtually all mathematical systems, and many of them think that sets should be basic to our system of instruction in the elementary and secondary schools. Thus, it becomes imperative that we explore the idea of sets.

A set is essentially a collection of things. Associated with a set is a rule which enables one to tell whether a given element belongs to the set or not. Without such a rule the idea of set might be vague. You can think of all the things on your desk as a set of things. This set of things may consist of: pencil, eraser, paper clip, pipe, and pen. In order to

indicate that you are to think of these things as a whole—a set—we make use of braces. The set is symbolized by '{pencil, eraser, typewriter, pen, paper clip, pipe}'. The '{ }' indicate that you are to think of this collection as a set of things and not as individual objects. But where is the rule which tells us whether a given object should be an element (a member) of the set? In this case the rule is suggested by 'the things on your desk'. If an object is on your desk it is in the set; if it is not on your desk it is not in the set.

The above set might appear in an elementary arithmetic class as a set of objects to be counted. In more advanced classes in the elementary school, the junior high school, and the senior high school, sets of numbers appear naturally. Let us see how this could happen.

Suppose that a class decides to put on a benefit show. They plan to have a 35 cent admission charge. (Note that the class itself is a set of individuals; namely, {Mary, Jane, Joe, $\cdots$, Sam}. The students are elements of the set, and the rule is 'Is a member of the sixth grade'.) In order to help the boys and girls who will sell tickets at the door the class makes a table such as illustrated in Table 1.

TABLE 1

Number of tickets	Cost (in dollars)
1	.35
2	.70
3	1.05
4	1.40
5	1.75
6	2.10
7	2.45
8	2 80

Since the class decides that it is unlikely that anyone will buy more than eight tickets, they don't extend the table any further than shown.

Now in this situation we see two sets of numbers. The one set we will call 'D'; $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and the other '$R$'; $R = \{.35, .70, 1.05, 1.40, 1.75, 2.10, 2.45, 2.80\}$. Note that D and R each have eight elements and eight elements only, and that the elements in D and R are paired as shown in Table 1. The element .70 of R is paired with the element 2 of D. This can be indicated by writing '(2, .70)'.

Let us look more closely at D. It has eight members. The number 2 is a member of D, or in symbols '$2 \in D$'. But 9 is not a member of D; in symbols '$9 \notin D$'. Similarly, $4.5 \notin D$ and $50 \notin D$. These and other numbers are excluded by the nature of the physical situation. We can de-

scribe D as the set of all whole numbers less than 9. Similarly, R contains only certain numbers and not others. It might be described as the first eight multiples of .35.

Some children might think that a graph would serve the purpose in this situation as well as a table. In such a case it would be wise to make a graph. This is given in Figure 1.

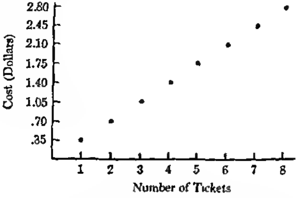


FIG. 1

You should notice that the line has not been drawn. The pupils should notice this also. The line is not drawn because there are only eight number pairs to plot. These eight number pairs make up a set which we call 'F'. $F = \{(1, .35), (2, .70), (3, 1.05), (4, 1.40), (5, 1.75), (6, 2.10), (7, 2.45), (8, 2.80)\}$.

Certainly, when teaching this, you would want to raise the question about other ways to show how the elements of two sets of numbers are paired. The pupils will have noticed that the numbers in the second column in the table are respectively the products of the correspondiag numbers in the first column and .35. Indeed, the table was constructed by carrying out such multiplications or else by successive additions of .35, which is another way of accomplishing the same result. Accordingly, it is natural to formulate a rule: 'The second number in each pair is the first times .35'. We now can easily formulate a rule that defines F: The first number in each pair must belong to D, and the corresponding second number must be the first number times .35. In other words, F = {all number pairs whose first members belong to D and whose second members are their first members times .35}.

The rule defining F can be stated in words, but such statements are cumbersome. Since the student has already had some experience with formulas, what could be more natural than to replace 'the second number must be the first number times .35' by '$y = .35x$'? Here, it is under-

stood that we may replace the letter 'x' by a name of any member of D and so obtain an equation giving y, the corresponding member of R. For example, if we replace 'x' by '2', we get $y = (.35) \times 2 = .70$. Knowing D and the equation we can find all the number pairs and so determine F.

From the illustration it is now possible to abstract the idea of a variable as it is used in mathematics. First, there is a set of numbers, D, which may be an arbitrary set; for example, $\{0, 4, 19, 456, 1980\}$ or all positive integers. Or it may be a set that is dictated by the requirements of some physical situation, as in the case just discussed or in a situation in which yards of cloth are purchased. In the last case, the numbers in the set under consideration must be positive real numbers not greater than the maximum number of yards for sale.

Now we may wish to talk about an arbitrary member of the set D. In order to do this we use a symbol, usually a letter of the alphabet, to stand for any member of the set. It now becomes possible to make sentences about any member of D. To illustrate, for the set D considered in the ticket problem above, we let 'x' stand for a name of a member of D. Then we can say that $x > 0$, $x \leq 8$, $x \in D$, and so on. Similarly, we can let 'y' stand for a name of a member of R. Then '$y = .35x$' is a sentence that must be satisfied by a pair (x, y) for the pair to belong to F. In fact, $F = \{$the set of (x, y) such that $x \in D$ and $y = .35x\}$; that is, $F = \{(x, y)$ such that $x \in D$ and $y = .35x\}$.

Let us notice several things about 'x' and 'y' in the examples above. First, they are symbols, letters of the alphabet. As anyone can plainly see, they are not numbers. They simply serve as placeholders, that is, they occupy places which may be occupied by other symbols, particularly, number symbols. The symbols that can be substituted are names of members of some set. We describe this by saying that 'x' and 'y' stand for the names of the elements of the sets D and R respectively. We call members of these sets values of 'x' and 'y'. We have emphasized that 'x' and 'y' are symbols by using single quotation marks. Thus, when we say that 'x' is a letter of the alphabet, we are talking about the letter 'x'. When we say that in a set D, $x \leq 8$, we are talking about the members of D. What we mean here is that the sentence '$x \leq 8$' yields a statement which is true if we substitute for 'x' a name of any member of D. We call 'x' and 'y' variables. Roughly speaking, a variable is a symbol for which one substitutes names for some objects, usually a number in algebra. A variable is always associated with a set of objects whose names can be substituted for it. These objects are called values of the variable.

We illustrate with a simple example involving only one variable. Let $U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Let '$x$' be a variable whose range is U; that is, the name of any member of U may replace 'x'. Then consider the sentence '$4x - 1 > 10$'. Substituting '0' for 'x' we get $4 \times 0 - 1 > 10$ which is false, and substituting '4' for 'x', we get $4 \times 4 - 1 > 10$ which is true. In this way the sentence '$4x - 1 > 10$' can be used as a set-builder to find a subset of U whose members satisfy it. By trial we can find that the set of numbers that belong to U and satisfy $4x - 1 > 10$ is $\{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Pedagogical Implications. The definition above describes a variable in terms of how it is used. By way of contrast consider the following definition still in general use. *A variable is a quantity that varies.* The pupil may well be mystified by this definition, since it is hard to see how the letter 'x' can *vary* or how a letter can be a *quantity*. In our examples, 'x' was merely a symbol that held a place in a sentence until we got ready to replace it with the name for some object (number).

To help emphasize this thought let us take another illustration. Consider the set $N = \{$all positive and negative integers and zero$\}$. Let 'x' be the placeholder for any member of N. Suppose we write the sentence $5x - 1 > 24$. As yet we have said nothing that can be reacted to as true or false. The sentence $5x - 1 > 24$ is much like the sentence 'It is a book'. In each case we cannot tell whether the sentence is true or false until, in the former, the number name to replace 'x' is specified and in the latter the antecedent of 'it' is specified. However, if we do replace '-5' by 'x' we see that the statement $5 \cdot (-5) - 1 > 24$ is false. Our next replacement may be '1000'. In this case $5 \cdot 1000 - 1 > 24$ is a true statement. Note that what we substitute for a variable are also symbols (e.g., numerals), though they are names of quite definite things (e.g., numbers). Such symbols, that stand for specific and unique objects, are called *constants*. Thus all numerals are constants. The numeral '2' is a name of (stands for) the number 2. Other numerals which are names for the same number are 'II', '$4 - 2$', and '$8/4$'. Substituting constants for variables is one of the important operations with which the student is familiar before he reaches high school.

You now see that the variable did not vary and cannot vary in the sense the word 'vary' is used ordinarily. The 'x' is only a placeholder in a mathematical sentence. It is a mistake to say that a variable *varies just like temperature varies*, because 'variable' is not used in that sense in mathematics. In mathematical terms we can describe the physical situation in which temperature varies as follows: If we measure the temperature at different times, we get a set of number pairs. In these

pairs the first component is, say, the time, and the second component is the corresponding temperature. We may use variables and say that a temperature, T, corresponds to each time, t. Then the variable 'T' takes different values at different times, but the variable itself (the letter 'T') does not vary, even though we can use it effectively to describe the physical changes in temperature.

This situation affords an excellent chance for you to teach vocabulary and language structure in a mathematics class. By clarifying the meanings of frequently used words you will save the student much confusion and speed his mathematical progress. To illustrate, let us examine a frequently used method for introducing the formula for the area of a rectangle. In the sixth or seventh grade, by one means or another, the pupil is led to the generalization that the 'area equals the length times the width'. This expression is shortened to: Area = Length × Width. The next step is the simple, in fact, too simple, step $A = L \times W$, where 'A' is the abbreviation for 'Area', 'W' for 'Width', and 'L' for 'Length'.

The pupil is accustomed to multiplying numbers. Now we appear to be asking him to multiply words and letters. Of course, in this context both words and letters are being used as variables (placeholders) whose role is to occupy a position in which the pupil is to substitute appropriate numerals. But this idea is frequently not emphasized.

Is it any wonder that children find mathematics hard? If they are not let in on the secret of how symbols are used in mathematics from the beginning it must be difficult to find out what it is all about. There is a great need for a re-evaluation of the methods used to introduce children to the use of variables in mathematics. The illustration used at the outset of this chapter would seem to be simpler than the formula for the area of a rectangle. It would serve more appropriately as a *first contact* with algebra than a formula for area.

Other Names for Some Uses of Variables. Variables appear in mathematics under different names. One of these is 'unknown'. For example, when we ask the student to solve the equation '$3x - 2 = 0$', we call 'x' an unknown. But we are really asking him to find the value(s) of the variable 'x' whose names make true statements when substituted for 'x' in '$3x - 2 = 0$'. So we see that an unknown is just a variable that appears when we solve equations.

Another alias is 'parameter'. Suppose we ask the student to solve the equation $2x - 2 = c$. Sometimes 'c' is described as a constant. Obviously, it is not a name of a number, and so is not a constant. It is more appropriately called a parameter. Clearly it is a variable, since we can substitute numerals for it. When we ask the student to solve $3x - 2 = c$, we expect him to write '$x = (c + 2)/3$'. Now this is the

formula that gives the solution to the equation for any value of 'c'. That is, we can substitute a numeral for the variable 'c' in both $3x - 2 = c$ and $x = (c + 2)/3$. Certainly then, 'c' plays the role of a variable, since it is a symbol for which we can substitute appropriate numerals. Generally, a variable for which we do not want to solve or which plays some auxiliary role is called a 'parameter' to distinguish it from the variables on which our attention is focused primarily.

Another example of the use of a parameter is in graphing an equation such as $x + y = c$. Actually, we cannot graph it until we substitute for 'c'. This fact alone makes clear that 'c' is not a constant. On the contrary, 'c' is a variable for which we substitute a name for any real number. When we do substitute we get a straight line which is a geometric picture of the set of pairs (x, y) that satisfy the equation. For example: $x + y = 2$ is satisfied by $(1, 1)$, $(2, 0)$, $(0, 2)$, $\cdots$. All of the solutions make up a set of pairs whose graph is a straight line. For each value of the parameter 'c' we get a set of number pairs whose graph is a line. Corresponding to the set of all choices for 'c' we have a family (set) of straight lines; so again, 'c' plays the role of a variable in a special way and is distinguished by the name 'parameter'.

RELATIONS

Many times in life you must use sets of number pairs or even sets of number triples, or quadruples. In fact, the number of times a day that one needs single numbers is not great. When the situation involves nothing more than counting, then a single number is sufficient to report or record the result. The number of eggs in a basket and the number of dollars in your purse are examples of such situations. Even here more than one number may be involved if the time or place of counting must be recorded.

In contrast, consider the items you buy at a grocery store. Here, number triples are usually involved. You know the price per item and the number of items. This pair of numbers enables you to find a third number which you interpret as the cost of the purchase. In reality you use a number triple. This situation prevails in the illustration used at the beginning of the chapter. The cost per ticket was not recorded because it was \$.35. You could almost ignore it and pay attention to only such number pairs as (3, 1.05). So it is not too surprising to be told that the study of number pairs, triples, and so on, is very useful as a means of recording or studying events in the world.

Look again at the table the fifth grade class made to help ticket sellers at their benefit show (Table I).

You have already noted that two sets of numbers were involved and

that the numbers in each of these two sets were put in one-to-one correspondence by the way their names were written in the table. This one-to-one correspondence was also given by means of the rule $y = .35x$. It was important to note too the range of the variable 'x' and the range of the variable 'y'. 'x' ranges over the set $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ and 'y' ranges over the set $R = \{.35, .70, 1.05, 1.40, 1.75, 2.10, 2.45, 2.80\}$.

There is still another idea buried in this set of number pairs which needs to be brought out in any study of elementary algebra. Before identifying this idea precisely let us take a look at several instances in which sets of number pairs have been plotted.

The set of number pairs $F = \{(1, .35), (2, .70), (3, 1.05), (4, 1.40), (5, 1.75), (6, 2.10), (7, 2.45), (8, 2.80)\}$ used by the fifth grade at their benefit show is graphed as in Figure 2.

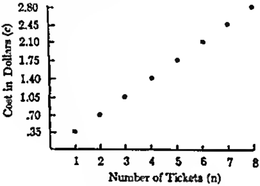


FIG. 2

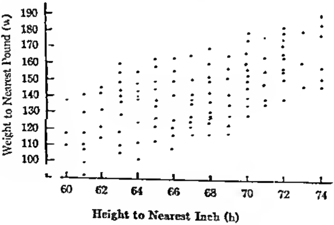


FIG. 3

The set of number pairs $\{(h, w)\}$ obtained by measuring the height, h, (nearest inch) and weight, w, (nearest pound) of the members of a high school class is plotted in Figure 3.

Figure 4 displays the set of pairs $\{(n, A)\}$ representing the amount, A, of money Mr. Brown has in his checking account ($100) after n years, assuming he forgot about his account entirely and received no interest during a five year period.

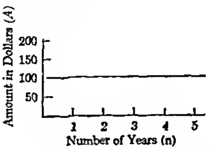


FIG. 4

The set of number pairs (x, y) given by the formula '$x + y = 6$' (where the range of 'x' and 'y' is the set of all real numbers) is graphed in Figure 5.

Figure 6 shows the set of number pairs $\{(n, c)\}$ which would be

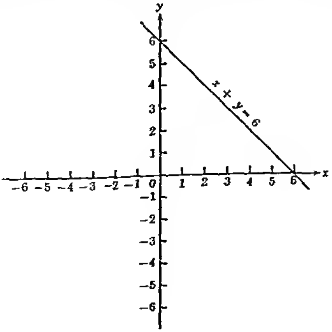


FIG. 5

obtained if postage were 3¢ for a letter weighing an ounce or less, 6¢ for a letter weighing more than one ounce but less than two, etc. The dot at the right end of each segment indicates that the right end point of each one (but not the left end point) is on the graph. (See page 108.)

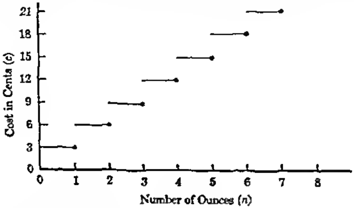


FIG. 6

The shaded area in Figure 7 represents the set of number pairs $\{(x, y)$ such that $x + y > 6\}$ (Note that the boundary line is not included!)

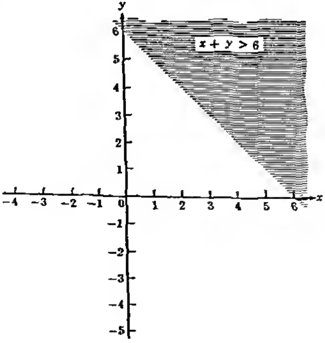


FIG. 7

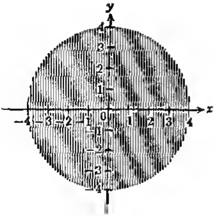


FIG. 8

while the shaded area in Figure 8 represents the set of pairs $\{(x, y)$ such that $x^2 + y^2 < 16\}$ (Note that the circular boundary is not included!). The graph of $\{(x, y)$ such that $y = x^2\}$ is shown in Figure 9. The graph of $\{(x, y)$ such that $y^2 = x\}$ is shown in Figure 10.

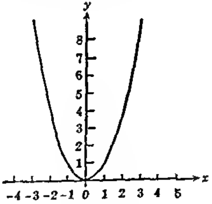


FIG. 9

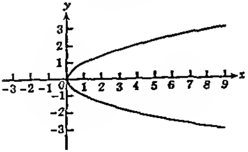


FIG. 10

Now let us examine these graphs and note some outstanding characteristics and outstanding differences.

1. In each case we have a set of number pairs whose graph is a set of points in the plane.

2. In each case there are two sets of numbers, the set of first components and the set of second components of the number pairs. The numbers in these sets are paired in a manner defined by the rule and visualized in the graph.

3. In some cases a member of the first set is paired with only one member of the second set (in Figs. 5 and 9, for example). In other cases several members of the second set are paired with the same first member (in Figs. 7 and 10, for example). It might be interesting for you to check all the examples and assign them to one of the above categories.

Relation Defined. The important thing for the secondary school student is to become familiar with the various ways to define relations —pairs, formulas, tables, graphs, and verbal descriptions. This is possible without defining in general what a relation is. However, the question "What is a relation?" is a natural one. It calls upon us to point to something and say, "That is the relation." Clearly the formula is not satisfactory, since many different formulas yield the same relation. On the other hand, the set of pairs itself seems a satisfactory choice, since if we know this set we know everything about the relation, and conversely. Accordingly, we say that a relation is a set of ordered pairs. Hence, any set of ordered pairs is a relation. If we are given a set of ordered pairs, we may describe it graphically or by means of a formula, a rule, a sentence, or a table. But the relation is not the linguistic or pictorial devices used to talk about it. The relation is the set of ordered pairs.

In the example of tickets for the benefit show pictured in Figure 1 on page 69 the relation is given by $F = \{(1, .35), (2, .70), \cdots, (8, 2.80)\} = \{(x, y)$ such that $x \in D$ and $y = .35x\}$. Here '$y = .35x$' is a formula defining the relation. The set of first components is $D = \{1, 2, \cdots, 8\}$. We call the set of first elements of a relation its domain. Hence, D is the domain of F. The set of second components of F is $R = \{.35, .70, \cdots, 2.80\}$. We call the set of second elements of a relation its range. Hence, R is the range of F.

One of the advantages of this point of view is that in elementary instruction we can identify the graph as the picture of the relation (that is, the picture of the set of ordered pairs), and this conforms to familiar ways of talking, thinking, and visualizing. The formula is then treated as a powerful computational device for talking about the graph and/or

the set of pairs. In each of the examples given in this chapter it is easy to identify the relation itself, its domain, its range, its graph, and one or more formulas that define it. For example for the very first relation mentioned (see page 66), the domain D is the set of hourly readings from 8 a.m. to 4 p.m. inclusive, the range R is a set of temperatures, the graph is easily sketched, and the relation F may be defined by $F = \{(x, y)$ such that $x \in D$ and $y =$ the temperature at the time $x\}$. In the second example (see page 66), the domain is a set of five states, the range a set of populations, and a defining rule is 'y is the population of x'.

To repeat, a relation F is a set of ordered pairs, with a domain, D, consisting of the first components, a range, R, consisting of the second components, and a rule (formula, sentence, condition), say '$r(x, y)$', by means of which we can define the relation by $F = \{(x, y)$ such that $x \in D$, $y \in R$, and $r(x, y)\}$. For example, in Figure 7, page 76, the rule is '$x + y > 6$'.

When defining a relation the words 'such that' are abbreviated in some way, the most common being the stroke'$\,|\,$'. Then the relation of Figure 9 is $\{(x, y) \mid x$ is a real number and $y = x^2\}$.

Important Properties of Relations. In the preceding paragraph we have defined a relation as a set of ordered number pairs. In what follows we use the letter 'R' to designate a relation. Thus, $R = \{(x, y) \mid x$ is a real number and $y = x^2\}$ is a relation.

It is common to write 'xRy' for $(x, y) \in R$. This notation is convenient at times. Daily language often prefers the 'xRy' designation rather than the set designation as given above. For example, 'x is the father of y' can be written as 'xRy' where R designates 'is the father of'. Note that 'is the father of' also defines a set of pairs. Thus, the series Mr. Smith R John; Mr. Smith R Susan; Mr. Smith R Fanny; Mr. Smith R Joe, where R designates 'is the father of' is more conveniently designated by $R = \{$(Mr. S, John), (Mr. S, Susan), (Mr. S, Fanny), (Mr. S, Joe)$\}$.

Many relations can be expressed via the 'xRy' symbolism. Here are only a few. (1) Let 'R' mean 'is greater than', then 'xRy' means $x > y$. (2) Let 'R' mean 'is taller than', then 'xRy' means 'x is taller than y'. You can supply many others. In each case it is instructive to set up the relation as a set of ordered pairs, and also in the 'xRy' way where the domain of 'x' and the range of 'y' are specified.

Now some relations have three especially important properties. The most common of such relations is that designated by '$=$'.

PROPERTY 1. *If xRx for all elements of the domain of R, then R is said to be reflexive.*

Not all relations are reflexive. Obviously, $x > x$ is not true for any

numerical replacement of 'x', but $x = x$ is true for all such replacements. Such relations as 'is the sister of' are not reflexive but 'as heavy as' is a reflexive relation.

PROPERTY 2. *If xRy then yRx for all x in the domain of R and all y in the range of R, then the relation is said to be symmetric.*

Equals is a symmetric relation, for if $x = y$ then $y = x$. On the other hand 'is the brother of' is not symmetric.

PROPERTY 3. *If xRy and yRz, then xRz. In this case R is said to be a transitive relation.*

Equals is a transitive relation, for if $x = y$ and $y = z$, then $x = z$. *Is the brother of* is not transitive; neither is the relation *likes* transitive but *is greater than* is transitive.

Relations which are symmetric, transitive, and reflexive occur frequently in mathematics. They also occur in nonmathematical situations. Consider the relation *goes to the same school as*, and let's apply it to all children who are ten years old in a given city at a given time. Designate this relation by 'R'. Now xRy, where 'x' and 'y' are placeholders for the names of children in the city, partitions all the children who are ten years old into groups (assume all the children go to a school), each group going to a particular school. This happens because R is an equivalence relation, that is, it is symmetric, reflexive and transitive. A little thought will show that such nonequivalence relations as *is less than* and *is the sister of* will not partition a set into mutually exclusive groups.

Some Uses of Equivalence Relations. We have already seen that equals is an equivalence relation in the set of integers. If we define two fractions a/b and c/d to be equal provided $ad = bc$ ($a/b = c/d$ if and only if $ad = bc$), then $=$ is an equivalence relation over the set of fractions. Let us convince ourselves that this is so. Since $2/3 = 2/3$ the equals relation is reflexive. If $2/3 = 4/6$, then $4/6 = 2/3$, so $=$ is symmetric. If $2/3 = 4/6$ and $4/6 = 8/12$, then $2/3 = 8/12$, so $=$ is transitive.

An equivalence relation partitions the set over which it is applied into disjoint subsets. Let us apply this principle to the fractions of arithmetic. We will select a few fractions and see what other names we can find for them. All names for the same fraction will be put into one set.

Start with $2/3$. What other names can we find for it? Obviously, '$4/6$', '$6/9$', '$8/12$', '$20/30$', '$22/33$', and so on. Now form the set

$$A = \left\{\frac{2}{3}, \frac{4}{6}, \frac{6}{9}, \frac{8}{12}, \frac{20}{30}, \cdots \frac{2n}{3n} \cdots\right\}.*$$

---

* The rational number 2/3 can now be defined as the set of all pairs a/b such that $2/3 = a/b$. See Edmund Landau: *Foundations of Analysis* (tr. by F. Steinhardt). New York: Chelsea Publishing Co., 1951.

In the same way form the set

$$B = \left\{\frac{1}{2}, \frac{2}{4}, \frac{3}{6}, \frac{4}{8}, \cdots \frac{n}{2n} \cdots\right\}$$

and the set

$$C = \left\{\frac{5}{6}, \frac{10}{12}, \frac{15}{18}, \cdots \frac{5n}{6n} \cdots\right\}.$$

If this process were to be continued *ad infinitum* you would find that no symbol occurs in two different sets. In other words, the set of fractions can be partitioned into mutually exclusive sets by means of the equivalence relation $=$. This is not true of such a relation as $>$. You may wish to show that this statement is true.

How are these ideas used in arithmetic? The pupil is faced with the task of adding $\frac{1}{2}$ and $\frac{2}{3}$. As it stands the problem is inconvenient. So he selects another member of B ($\frac{3}{6}$) and another member of A ($\frac{4}{6}$). He knows as a result of previous instruction that '$\frac{3}{6}$' is just another name for $\frac{1}{2}$ so it can be used instead of '$\frac{1}{2}$'. Now the thing he should see is that any number pair in B plus any number pair in A always results in a number pair which is an element of C. In other words, the pupil has a great deal of liberty as to how he can go about adding $\frac{1}{2}$ and $\frac{2}{3}$. He knows he'll never go wrong if he substitutes the names from one equivalence class for a name in that same class.

FUNCTIONS

The mathematical concept of relation as described on the preceding pages is fairly close to the idea of relation as used outside mathematics. In contrast, the meaning of function in mathematics is practically independent of the ordinary usage of this word in such expressions as "The function of the heart is to circulate the blood." However, the modern mathematical concept of function is linked naturally to the historic use of the word 'function' by mathematicians and scientists. Historically, 'function' has been used to indicate conditions under which one thing determines another. For example, the physicist may say that *distance is a function of time* meaning that if we know the time we can find the distance. Now in some of the relations considered in this chapter, a knowledge of the first component of a pair belonging to a known relation tells us what the second component must be. This was true in the case of the states and population, for example, since each state has a unique population. We call such relations 'functions'.

We say that a function is a relation in which each element of the domain belongs to only one pair. This means that each first element is

paired with only one second element. It means also that a rule defining the relation must be such that if a value of 'x' is given in the domain, there is only one corresponding value for 'y'. Thus the relation in Figure 7 (page 76) is not a function since for a given value of 'x', $x + y > 6$ has many solutions.

The graph of a function reflects in a very simple way the fact that a unique second component corresponds to each first component. Any line drawn parallel to the y-axis crosses the graph of a function only once. Figure 9 (page 77) illustrates this very clearly. In Figure 10, a vertical line, two units to the right of the origin, intersects the graph of $y^2 = x$ twice. Hence, the relation $\{(x, y) \mid y^2 = x\}$ is not a function. But the relation $\{(x, y) \mid y = x^2\}$ shown in Figure 9 is a function, since any line parallel to the y-axis intersects the graph, if at all, only once. You may be interested in applying the above criteria for determining which of the relations in Figures 2 to 8 are also functions.

You should see from the definitions given for function and relation that a set of pairs is always a relation but that it may be a special kind of relation; namely, one in which a unique second element is paired with each of its first elements. Hence, a function is a relation but a relation is not necessarily a function.

CARTESIAN PRODUCTS

When we graph a set of pairs (x, y) that satisfy a sentence such as $y = x^2$, we are selecting a subset of all the points of the plane to picture the relation. In other words, the universe in which we are setting up the locus of points such that $y = x^2$ is the set of all pairs whose components are real numbers. Let 'L' be the name for the set of all real numbers; the set of all points in the plane from which we select our graph is $\{(x, y) \mid x \in L \text{ and } y \in L\}$. It is convenient to have a short name for this set which is the entire plane. We call it the Cartesian product (after René Descartes) of L and L and symbolize it by '$L \times L$'.

More generally, if A and B are any two sets whatsoever, $A \times B$ is the set of all pairs that can be formed by selecting a first component from A and a second from B.

Suppose we are discussing $A = \{1, 2, 3\}$. Then $A \times A$ is the set of all those number pairs whose components are chosen from A. In symbols, $A \times A = \{(x, y) \mid x \in A \text{ and } y \in A\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$. The graph of the set $A \times A$ on a rectangular coordinate is shown in Figure 11.

Let $A = \{1, 2\}$ and $B = \{3, 4, 5\}$. Then the Cartesian product $A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\} = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}$. The graph is shown in Figure 12.

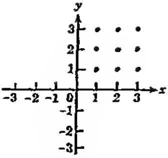


FIG. 11

To fix the idea you may be interested in forming and graphing other Cartesian products. Let L be the set of all real numbers, R the set of all rational numbers, N the set of all positive and negative integers, P the set of all positive integers. On rectangular axes plot some members of the following sets: (1) $P \times P$; (2) $N \times P$; (3) $L \times P$: (4) $L \times L$; (5) $N \times R$; and (6) $R \times R$.

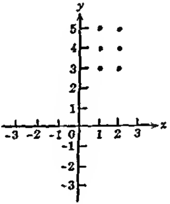


FIG. 12

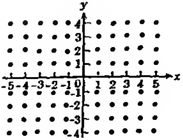


FIG. 13

paired with only one second element. It means also that a rule defining the relation must be such that if a value of 'x' is given in the domain, there is only one corresponding value for 'y'. Thus the relation in Figure 7 (page 76) is not a function since for a given value of 'x', $x + y > 6$ has many solutions.

The graph of a function reflects in a very simple way the fact that a unique second component corresponds to each first component. Any line drawn parallel to the y-axis crosses the graph of a function only once. Figure 9 (page 77) illustrates this very clearly. In Figure 10, a vertical line, two units to the right of the origin, intersects the graph of $y^2 = x$ twice. Hence, the relation $\{(x, y) \mid y^2 = x\}$ is not a function. But the relation $\{(x, y) \mid y = x^2\}$ shown in Figure 9 is a function, since any line parallel to the y-axis intersects the graph, if at all, only once. You may be interested in applying the above criteria for determining which of the relations in Figures 2 to 8 are also functions.

You should see from the definitions given for function and relation that a set of pairs is always a relation but that it may be a special kind of relation; namely, one in which a unique second element is paired with each of its first elements. Hence, a function is a relation but a relation is not necessarily a function.

CARTESIAN PRODUCTS

When we graph a set of pairs (x, y) that satisfy a sentence such as $y = x^2$, we are selecting a subset of all the points of the plane to picture the relation. In other words, the universe in which we are setting up the locus of points such that $y = x^2$ is the set of all pairs whose components are real numbers. Let 'L' be the name for the set of all real numbers; the set of all points in the plane from which we select our graph is $\{(x, y) \mid x \in L \text{ and } y \in L\}$. It is convenient to have a short name for this set which is the entire plane. We call it the Cartesian product (after René Descartes) of L and L and symbolize it by '$L \times L$'.

More generally, if A and B are any two sets whatsoever, $A \times B$ is the set of all pairs that can be formed by selecting a first component from A and a second from B.

Suppose we are discussing $A = \{1, 2, 3\}$. Then $A \times A$ is the set of all those number pairs whose components are chosen from A. In symbols, $A \times A = \{(x, y) \mid x \in A \text{ and } y \in A\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$. The graph of the set $A \times A$ on a rectangular coordinate is shown in Figure 11.

Let $A = \{1, 2\}$ and $B = \{3, 4, 5\}$. Then the Cartesian product $A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\} = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}$. The graph is shown in Figure 12.

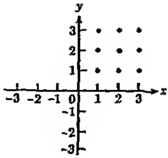


FIG. 11

To fix the idea you may be interested in forming and graphing other Cartesian products. Let L be the set of all real numbers, R the set of all rational numbers, N the set of all positive and negative integers, P the set of all positive integers. On rectangular axes plot some members of the following sets: (1) $P \times P$; (2) $N \times P$; (3) $L \times P$: (4) $L \times L$; (5) $N \times R$; and (6) $R \times R$.

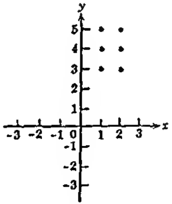


FIG. 12

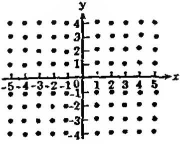


FIG. 13

Now let's think of $N \times N$ as the universe in which we are working. Figure 13 shows the graph of $N \times N$ for a portion of the plane.

If we wish to consider the set of points for which $y = x^2$ in this universe, the next step consists of selecting a subset of $N \times N$ according to the rule '$y = x^2$'. This subset is $S = \{(x, y) \mid x \in N, y \in N, \text{and } y = x^2\}$. A few of the elements of S are $(1, 1)$, $(-1, 1)$, $(2, 4)$, $(-2, 4)$, $(3, 9)$, and $(-3, 9)$. Figure 14 shows part of S as a subset of $N \times N$. Compare this with Figure 9. It shows the graph of the same relation in $L \times L$.

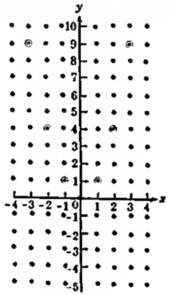


FIG. 14

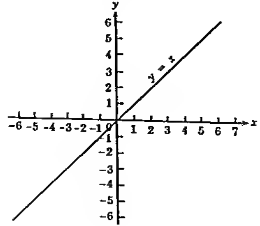


FIG. 15

If the universe in which we are working is $L \times L$ then the graph $L \times L$ is the whole plane. A subset of $L \times L$ may be selected according to the rule '$y = x$'. This subset is $K = \{(x, y) \mid x \in L, y \in L, \text{and } y = x\}$. If we plot the elements of K there appears the familiar straight line as in Figure 15.

On the other hand, if the universe in which we are working is $N \times N$, then the subset $K' = \{(x, y) \mid x \in N, y \in N, \text{and } y = x\}$ gives rise to the graph in Figure 16.

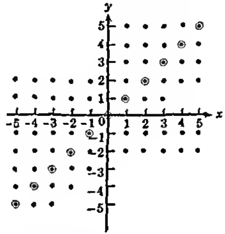


FIG. 16

The basic idea of graphing is very simple. One sets up a one-to-one correspondence between the totality of points in a universe and the totality of number pairs. This totality of pairs is a Cartesian product. Then a subset of the universe is identified by a condition that selects certain pairs. The result is called a graph of the given set and also the graph of the condition.

We can now give a better definition of relation and function.

DEFINITION: *A relation over a set A is a subset of $A \times A$. If 'R' denotes the relation and $(a, b) \in R$, then at times we write aRb.*

DEFINITION: *A relation which has only one second element paired with each first element of the number pairs is called a 'function'.*

POINTS OF VIEW PERTAINING TO RELATIONS AND FUNCTIONS

The definitions of a relation as a set of ordered pairs and of a function as a special kind of relation are enlightening and useful in many ways. But it is evident that a great deal can be done with particular relations and functions before the student reaches the stage at which this point of view is necessary. Indeed, he must become familiar with many specific

cases before he is ready to grasp the general idea. Functions and relations appear in many quite different situations, in many disguises, and under many different names. Hence, they can be treated from many points of view, each of which makes a contribution to a full understanding.

Often a relation is described as a correspondence. We think of the set of pairs as a set of corresponding numbers or objects. We think of the rule as defining the correspondence. We say that the rule defines the correspondence by telling which things correspond. If the relation is a function, then just one object in the range corresponds to each object in the domain. If F is a function, we may write '$F(x)$' as a name for the object corresponding to 'x'. Then the rule may be written in the form '$y = F(x)$'. For example, in the benefit ticket situation, $F(x) = .35x$.

This familiar functional notation has given rise to much misunderstanding. The practice of reading '$y = F(x)$' as *y is a function of x* is very confusing since y is not a function at all. This way of talking is related to such statements as *cost is a function of price*, which means that cost depends on price. But in such expressions 'function' is not being used in the mathematical sense. It is better to read '$y = F(x)$' as 'y equals F at x' or 'y = the object that corresponds to x'.

Another common way of speaking that sometimes causes confusion is to describe '$f(x)$' as *the value of the function corresponding to x*. This arises from confusing the formula with the function itself. In the function defined by $y = x^2$, $f(x) = x^2$. If we say that *y is a function of x* and identify 'y' and '$f(x)$' with the function itself, then it is natural to say that *y is the value of the function*. But the symbol '$f(x)$' has the advantage that it involves 'x' explicitly. This makes it convenient for indicating the *y corresponding to a particular x*. For example, $f(2)$ is the value corresponding to 2. This use of the functional notation is familiar and has lost none of its importance.

The functional notation can be used to write a more compact expression for a set of pairs that is a function. Since there is just one value of 'y' for each value of 'x', we may use '$f(x)$' to represent this y and define the function as the set of pairs $f = \{(x, f(x)) \mid x \in D\}$. For example, $\{(x, y)$ such that x is a real number and $y = x^2\} = \{(x, x^2) \mid x \text{ is real}\}$. We might even write just '$\{(x, x^2)\}$' leaving the domain to be understood.

Another very useful point of view arises from considering maps. In a map of a terrain, each point on the map corresponds to a point of the terrain, each map feature to a feature of the terrain. Here is a set of pairs in which the first member is a point on the ground, the second point on a map. Similar remarks apply to any plan, chart, or pictorial representation. The fact that maps are instances of functions gives a way of intro-

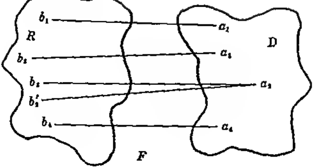


FIG. 17

ducing the ideas of sets, pairings, and relations in terms familiar to very young pupils.

But the above observation has much greater application. Every relation may be considered as a pairing of objects with their maps. As suggested in Figure 17, we may think of the domain D of the relation as a set that is mapped onto the range R. Note that in Figure 17 the point a_2 is mapped into two different points b_2 and b_2'. Hence, this mapping does not yield a function. When we speak of relations as mappings we speak of the objects in the range as the images of the objects in the domain. We also speak of object in the domain being carried into corresponding objects in the range. This terminology is often suggestive and gives a way to visualize a relation that is quite different from the familiar graphs. Obviously, it may be used in cases where a graph might not be helpful. For example, the son-parent relation defined by 'x is a son of y' may be visualized by thinking of the domain D consisting of all sons and the range T consisting of all parents. Clearly every son has two parents, so there are two y-values corresponding to each x-value. In Figure 17 we may think of b_2 and b_2' as the father and mother of a_2. Even where a graph is convenient, the mapping point of view may be informative. For example, the function defined by $y = x^2$ maps all real numbers into nonnegative real numbers, an important fact for the student to realize and one that can be easily visualized by thinking of the sets of real numbers and the set of nonnegative reals.

We have presented several different points of view from which functions (and relations) may be considered. We may describe them as sets of pairs, sets of points, tables, correspondences, or as mappings. We may emphasize the rule or we may concentrate attention on the set. In the historical development of mathematics these different aspects have been stressed to a varying degree at different times. The modern point of view is not contradictory to any of them, but unifies them all and

makes clear that they are all different ways of talking about the same thing—sets of ordered pairs.

One can almost hear the argument, "What's been good enough for mathematicians for over 100 years ought to be good enough for the next 100 years. Why change?" The reason lies, of course, "deep in the heart of mathematics."

For more than half a century mathematicians have been working on the foundations of their subject. They have asked questions about the key ideas of mathematics in their effort to get these ideas arranged in a logical sequence. They have been structuring mathematics. In the process, mathematicians have found that the idea of set is very basic, in fact, it is an idea on which much of mathematics can be built. This means that they are talking a *set language* more now than they did fifty, or even ten years ago. It further means that the definitions of fundamental mathematical ideas are stated in terms of sets or in terms which are readily reducible to sets.

This drive to reduce mathematics to its simplest and clearest terms is reflected in the definition of function and variable. The language is so chosen and the concept so organized that all ideas lead back to certain very elementary ideas about sets and elements of sets.

To illustrate the advantages of using *set language* let us try to formulate a definition of function in terms of the concept of correspondence. We may say that a function is a correspondence in which a unique object corresponds to each one of certain objects. But what is a correspondence? The definition is not very enlightening until we know what a correspondence is. It does not seem that 'correspondence' should be undefined in mathematics, yet it does not seem possible to define it except in terms of sets of pairs. And if we do this, we may as well define a relation in terms of sets of pairs and consider 'correspondence' as another term for talking about relations. By using the concept of a set as basic, we remove the mystery from our subject and define important concepts in simple terms more understandable to everyone.

It is not possible here to discuss all the different terminology used to talk about relations and functions. However, if the concept is understood it is easy to interpret properly most discussions found in mathematical and scientific writing. For example, one often finds a function defined as a correspondence in which to every value of one variable there corresponds one or more values of another variable. If just one value corresponds, the function is called single-valued, otherwise multiple-valued. Writers who use this terminology (and their number is decreasing) are calling functions 'single-valued functions' and other relations 'multiple-valued functions'.

FUNCTION AND RULE

In much of the work with functions, attention is concentrated on the rule that defines the set of pairs. Three typical situations arise.

Situation I occurs when a set of number pairs is given and we wish to find, if possible, a formula that enables us to pair the elements in the two sets. This kind of problem occurs frequently in statistics, where the given number pairs are the result of approximate observations. It occurs in such simple problems as finding the equation of a straight line on which we are given two points or a point and the slope. In such cases the function is defined without explicitly giving a formula, and the problem is to find the formula.

Situation II occurs when a formula is known, so that we have the function defined by the forms '$\{(x, f(x))\}$' or '$\{(x, y) \mid y = f(x)\}$' and we wish to calculate the second element in a pair whose first element is given. This is a matter of substitution in the formula. It is among the first algebraic manipulations that the pupil encounters.

Situation III occurs when a formula is given and we wish to calculate first elements corresponding to a second element. We say elements, since there may be more than one. For example, the function defined by '$y = x^2$' in the domain of real numbers is $\{(x, y) \mid x \text{ is real and } y = x^2\}$. We may ask, "How can we fill the blank in $(\_, 9)$ so that this number pair belongs to the function?" The answer is that either 3 or -3 will do. The problem from an algebraic point of view is that of solving $y = x^2$ for x when y is given. Thus the whole problem of the solution of equations may be viewed as an aspect of the study of functions and relations.

All three of these situations occur in the practical (and theoretical) applications of mathematics. Hence, it is important in the study of algebra that students learn to work with functions and relations in these three different situations. In common terminology, Situation I is usually called curve fitting or finding the equation of a locus. Situation II is usually called substitution and evaluation; it is introduced in all beginning algebra courses at an early date. Situation III is classified as equation solving.

INVERSE FUNCTIONS AND RELATIONS

In the study of sets of number pairs, it sometimes becomes convenient, or even necessary, to study the set of number pairs obtained by interchanging the first and second elements of each pair in the set. Thus, rather than study the set (function) $f = \{(x, x^2)\}$, it becomes convenient to study the relation $r = \{(x^2, x)\}$. The relation obtained by interchanging first and second elements in the pairs of a given relation is called

makes clear that they are all different ways of talking about the same thing—sets of ordered pairs.

One can almost hear the argument, "What's been good enough for mathematicians for over 100 years ought to be good enough for the next 100 years. Why change?" The reason lies, of course, "deep in the heart of mathematics."

For more than half a century mathematicians have been working on the foundations of their subject. They have asked questions about the key ideas of mathematics in their effort to get these ideas arranged in a logical sequence. They have been structuring mathematics. In the process, mathematicians have found that the idea of set is very basic, in fact, it is an idea on which much of mathematics can be built. This means that they are talking a *set language* more now than they did fifty, or even ten years ago. It further means that the definitions of fundamental mathematical ideas are stated in terms of sets or in terms which are readily reducible to sets.

This drive to reduce mathematics to its simplest and clearest terms is reflected in the definition of function and variable. The language is so chosen and the concept so organized that all ideas lead back to certain very elementary ideas about sets and elements of sets.

To illustrate the advantages of using *set language* let us try to formulate a definition of function in terms of the concept of correspondence. We may say that a function is a correspondence in which a unique object corresponds to each one of certain objects. But what is a correspondence? The definition is not very enlightening until we know what a correspondence is. It does not seem that 'correspondence' should be undefined in mathematics, yet it does not seem possible to define it except in terms of sets of pairs. And if we do this, we may as well define a relation in terms of sets of pairs and consider 'correspondence' as another term for talking about relations. By using the concept of a set as basic, we remove the mystery from our subject and define important concepts in simple terms more understandable to everyone.

It is not possible here to discuss all the different terminology used to talk about relations and functions. However, if the concept is understood it is easy to interpret properly most discussions found in mathematical and scientific writing. For example, one often finds a function defined as a correspondence in which to every value of one variable there corresponds one or more values of another variable. If just one value corresponds, the function is called single-valued, otherwise multiple-valued. Writers who use this terminology (and their number is decreasing) are calling functions 'single-valued functions' and other relations 'multiple-valued functions'.

FUNCTION AND RULE

In much of the work with functions, attention is concentrated on the rule that defines the set of pairs. Three typical situations arise.

Situation I occurs when a set of number pairs is given and we wish to find, if possible, a formula that enables us to pair the elements in the two sets. This kind of problem occurs frequently in statistics, where the given number pairs are the result of approximate observations. It occurs in such simple problems as finding the equation of a straight line on which we are given two points or a point and the slope. In such cases the function is defined without explicitly giving a formula, and the problem is to find the formula.

Situation II occurs when a formula is known, so that we have the function defined by the forms '$\{(x, f(x))\}$' or '$\{(x, y) \mid y = f(x)\}$' and we wish to calculate the second element in a pair whose first element is given. This is a matter of substitution in the formula. It is among the first algebraic manipulations that the pupil encounters.

Situation III occurs when a formula is given and we wish to calculate first elements corresponding to a second element. We say elements, since there may be more than one. For example, the function defined by '$y = x^2$' in the domain of real numbers is $\{(x, y) \mid x \text{ is real and } y = x^2\}$. We may ask, "How can we fill the blank in $(\_, 9)$ so that this number pair belongs to the function?" The answer is that either 3 or -3 will do. The problem from an algebraic point of view is that of solving $y = x^2$ for x when y is given. Thus the whole problem of the solution of equations may be viewed as an aspect of the study of functions and relations.

All three of these situations occur in the practical (and theoretical) applications of mathematics. Hence, it is important in the study of algebra that students learn to work with functions and relations in these three different situations. In common terminology, Situation I is usually called curve fitting or finding the equation of a locus. Situation II is usually called substitution and evaluation; it is introduced in all beginning algebra courses at an early date. Situation III is classified as equation solving.

INVERSE FUNCTIONS AND RELATIONS

In the study of sets of number pairs, it sometimes becomes convenient, or even necessary, to study the set of number pairs obtained by interchanging the first and second elements of each pair in the set. Thus, rather than study the set (function) $f = \{(x, x^2)\}$, it becomes convenient to study the relation $r = \{(x^2, x)\}$. The relation obtained by interchanging first and second elements in the pairs of a given relation is called

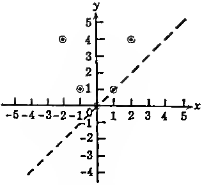


FIG. 18a. Function k.

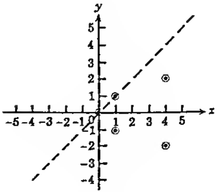


FIG. 18b. Relation r.

the inverse of the given relation. Let's see how this idea works by first studying a finite set (it's simpler) to get across the idea.

The function $k = \{(0, 0), (1, 1), (-1, 1), (2, 4), (-2, 4)\}$ has as its inverse $r = \{(0, 0), (1, 1), (1, -1), (4, 2), (4, -2)\}$. Note that the inverse of the function k is the relation r (r is not a function). The graphs shown in Figures 18a and 18b will help clarify this point.

In Figure 18b you can see that the condition that just one y-value corresponds to each x value is not fulfilled since $(1, 1)$ and $(1, -1)$ are elements of r.

Consider a second example. Let f' be the function $\{(1, 4), (2, 1), (3, 5), (4, 6)\}$. The inverse of f' is $r' = \{(4, 1), (1, 2), (5, 3), (6, 4)\}$. r' is a relation but it is also a function since it does not contain two pairs with identical first elements.

Graphically it is readily apparent that the point (b, a) is a reflection of the point (a, b) in the bisector of the 90° angle in the first quadrant. Since (b, a) belongs to the inverse relation if and only if (a, b) belongs to the original relation, graphing an inverse is simply a matter of reflecting the original graph in the line defined by '$y = x$'. For example, from Figure 18a to Figure 18b $(-2, 4)$ is reflected into $(4, -2)$, $(2, 4)$ into $(4, 2)$, and so on.

It is often of interest to inquire as to whether the inverse of a given function is also a function. A condition is quite readily written. The inverse of a function is also a function provided no two different pairs belonging to the original relation have the same second element. In other words, for both a function and its inverse to be functions, each first element must appear once and only once and each second element once and only once. Graphically, any horizontal line must cross the graph of the original function at most once. For example, the function of Figure

19 and its inverse are evidently both functions. On the other hand, the inverse of the function graphed in Figure 9 (page 77) is not a function, as is evident from the graph of this inverse in Figure 10 (page 77).

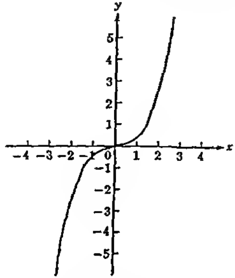


FIG. 19

The concept of inverse is closely related to situations of type III discussed above; namely, those in which we are given the formula defining a function and wish to find the first component (or components) corresponding to a given second component.

Since the inverse relation is obtained by interchanging components in the pairs, the formula defining the inverse is obtained by interchanging variables. If a function is defined by '$y = f(x)$', we are most often concerned with problems of type II, that is, given x to find y. If we are concerned with a problem of type III, that is, if $y = f(x)$, and we are given y to find x, we may view the problem as of type II by interchanging variables to get $x = f(y)$, that is, to get a formula defining the inverse function. Another way of describing this is to say that whether or not we consider a function or its inverse is just a matter of which set of objects we wish to consider as first components. Going from a function to its inverse interchanges the elements in each pair. But it is simply a different way of looking at the same situation.

In terms of the mapping idea, the inverse of a relation is the mapping that *goes the other way*, that maps the old range (now the domain of the inverse) into the old domain (now the range of the inverse). It is the reverse mapping.

In order to find the rule for the inverse one simply interchanges variables and solves for the other variable. If the solution is unique, we have a formula that defines the inverse function. For example: the inverse of the function defined by '$y = 2x - 6$' is obtained by solving $x = 2y - 6$ to get $y = (x + 6)/2$. The graphs of the two functions are shown in Figure 20.

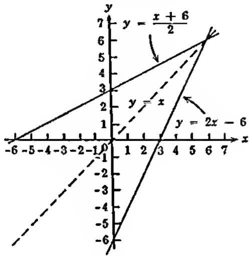


FIG. 20

Here it is readily observed that the one line is a reflection of the other line in the bisector of the 90° angle in Quadrants I and III.

In this example the inverse of a function is a function. By way of contrast consider the function defined by '$y = x^2$'. The inverse is defied by '$x = y^2$'. Solving we get $y = \sqrt{x}$ or $y = -\sqrt{x}$. For each value of 'x' we get two y-values in the inverse $\sqrt{x}$ and $-\sqrt{x}$, as illustrated in Figure 10. (Note that '$\sqrt{x}$' stands for just one number, the non-negative number whose square is x.)

It is possible to define one-to-one correspondence very simply in terms of the function and inverse concepts. If the inverse of a function is also a function we call it a one-to-one correspondence. This gives a precise characterization of the familiar process of establishing one-to-one correspondence by pairing off members of sets. Sometimes functions whose inverses are not functions are called many-one correspondences, and their inverses are called one-many correspondences.

UNIONS AND INTERSECTIONS OF RELATIONS

Since relations are sets, we can apply to them the techniques of set operations. First we recall the meaning of union and intersection. Con-

sider the set $A = \{1, 2\}$ and the set $B = \{5, 6, 7, 8\}$. The union of A and B is $A \cup B = \{1, 2, 5, 6, 7, 8\}$. The union of the two sets is simply the two sets *thrown in together* to make one set. This is the simplest of ideas, and is understood by even young children.

By *the intersection of two sets S and T*, we mean the set of elements which are common to both of the sets S and T. Thus the intersection of $S = \{0, 1, 2, 3, 4, 5\}$ and $T = \{4, 5, 6, 7, 8\}$ is the set $S \cap T = \{4, 5\}$. The intersection of the sets A and B of the previous paragraph has no members. We call a set with no members *the null set*, and we write $\{1, 2\} \cap \{5, 6, 7, 8\} = \emptyset$. Sets whose intersection is the null set are called disjoint sets. Young children are used to forming the union of disjoint sets. Indeed, the number of members in the union of two disjoint sets is the sum of the number of members in the individual sets. However, we can still form the union of two sets that are not disjoint. For example,

$$S \cup T = \{0, 1, 2, 3, 4, 5\} \cup \{4, 5, 6, 7, 8\} = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}.$$

In previous sections we have classified the set of number pairs $\{(x, y) \mid y^2 = x\}$ as a relation which is not a function. You will readily recall the graph of this set. It is shown in Figure 21. Let us call this relation 'r'.

Now consider the two functions $f = \{(x, y) \mid y = \sqrt{x}\}$ and $F = \{(x, y) \mid y = -\sqrt{x}\}$. The graphs of the function f and the function F are shown in Figures 22 and 23.

The figures (21, 22, and 23) indicate very clearly that the relation $r = \{(x, y) \mid y^2 = x\}$ is just the union of the function f and F, since if

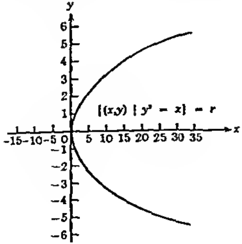


FIG. 21

we join the two graphs in Figures 22 and 23, we get the graph in Figure 21. In symbols, $r = f \cup F$.

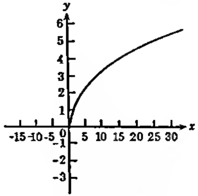


FIG. 22. Function f.

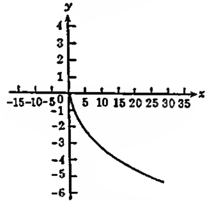


FIG. 23. Function F.

The reader can easily verify for himself that the relation

$$C = \{(x, y) \mid x^2 + y^2 = 16\}$$

is the union of the two functions $g = \{(x, y) \mid y = \sqrt{16 - x^2}\}$ and $G = \{(x, y) \mid y = -\sqrt{16 - x^2}\}$. Indeed, the graph of C is a circle of radius 4 with center at the origin. The graph of g is the upper half of this circle, and the graph G is the lower half. Evidently $C = g \cup G$.

Any relation can, in a similar way, be viewed as a union of functions. Of course, this can usually be done in several different ways, and more than two functions may be required.

A few more illustrations of how the union and intersection of two sets might occur in high school algebra may be helpful. Consider the set $S = \{(x, y) \mid x + y > 6\}$ and the set $T = \{(x, y) \mid x - y < 4\}$. The graphs of S and T are shown in Figure 24.

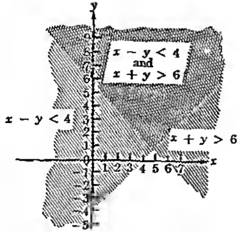


FIG. 24

The graph of the union of S and T ($S \cup T$) is the set of all points in any shaded region of the plane. The intersection of S and T, ($S \cap T$), is the set of all points in the doubly shaded region of the plane.

You now have a picture of $S \cup T = \{(x, y) \mid x + y > 6 \text{ or } x - y < 4\}$ (all the shaded area) and $S \cap T = \{(x, y) \mid x + y > 6 \text{ and } x - y < 4\}$ (the doubly shaded area). Notice that for the union the disjunction 'or' was used between the two conditions $x + y > 6$ and $x - y < 4$. In other words, the number pair (x, y) must be in S or in T or in both S and T if it is an element of $S \cup T$. For the intersection the conjunction 'and' was used. In this case the number pair (x, y) must be in both S and T if it is to be an element of $S \cap T$.

Of course, you are familiar with the usual solution of two conditions such as $x + y = 10$ and $x - y = -8$. The solution is $x = 1$ and $y = 9$. How would this familiar problem look in terms of the set terminology?

The sentences $x + y = 10$ and $x - y = -8$ are conditions that define relations. The condition $x + y = 10$ gives rise to the set

$$K = \{(x, y) \mid x + y = 10\},$$

and the condition $x - y = -8$ gives rise to the set

$$Q = \{(x, y) \mid x - y = -8\}.$$

The graphs of K and Q are the lines shown in Figure 25.

In other words, we have pictured all the number pairs that belong to K and the pairs that belong to Q, or we have graphed $K \cup Q$. $K \cup Q$ is the set of all number pairs which *graph into* points on either of the two lines. Now what is $K \cap Q$? It is the number pair associated with the

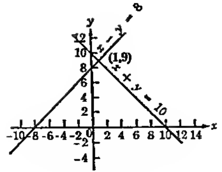


FIG. 25

intersection of the two lines in the graph. $K \cap Q = \{(1, 9)\}$. This number pair corresponds to the intersection of the two lines.

Pedagogical Implications. Why use set terminology and symbolism when solving simultaneous equations? You have no doubt experienced pupils solving two equations such as $x + 2y = 6$ and $x - 2y = 4$ by writing:

$$\begin{array}{rl} x + 2y = & 6 \\ x - 2y = & 4 \\ \hline 2x \quad = & 10 \\ x \quad = & 5 \end{array} \qquad \begin{array}{r} 5 + 2y = 6 \\ y = \frac{1}{2} \end{array}$$

Subsequent questioning showed that the pupil did not think of using 5 and $\frac{1}{2}$ to form a pair, $(5, \frac{1}{2})$. This is an essential part of the solution.

Observe the difference if we ask: Find the set $\{(x, y) \mid x + 2y = 6$ and $x - 2y = 4\}$. In this case the question asks explicitly for a set of pairs. Of course, if taught properly, the usual language also asks for pairs. But the set language helps the pupil understand what he is doing and how his algebra is related to the graph. Moreover, the set ideas may be used to solve simultaneous equations by graphical means or to find intersections before the pupil is facile with the mechanics of algebra.

By using the concepts of union and intersection, the teacher can show students how to write conditions defining quite elaborate sets of points in the plane. Experience shows that students find it thrilling to be able to define a set such as the interior of a square as the intersection of two sets. The student can easily verify that

$$\{(x, y) \mid 0 < x < 1\} \cap \{(x, y) \mid 0 < y < 1\}$$

is just the interior of the unit square resting on the positive x-axis with lower left corner at the origin. By graphing other examples and con-

structing examples of his own, the student will learn a great deal about equations and inequalities.

Generally, in high school work the universe of discourse is taken for granted. It is $L \times L$, although most points that students plot are in $R \times R$, that is, the points that have rational coordinates. But should the student always be working in $L \times L$? Should he not at times work in $P \times P$ or in still smaller Cartesian products? For example, candy bars cost 5 cents each. Assuming that no more than 12 bars will ever be purchased, use this information to construct a graph. From the physical situation one sees that two sets are involved.

$$C = \{0, 5, 10, 15, \cdots, 60\}$$

and $B = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. The universe in which we are working is $B \times C$ and we select a subset

$$\{(x, y) \mid x \in B, y \in C \text{ and } y = 5x\}.$$

The resulting graph imbedded in its universe is shown in Figure 26.

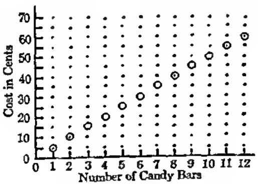


FIG. 26

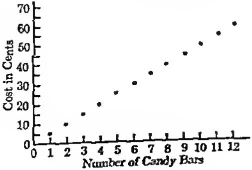


FIG. 27

This same graph if drawn by an eighth or ninth grader should look like that shown in Figure 27.

Too much of the graphing in junior high school is done without due consideration of the range of the variables involved. In the example above the range of the variable 'x' is B and the range of the variable 'y' is C. The physical situation tells us that in each case only a subset of the positive integers is under consideration. Hence, the graph is shown as a series of dots only. It is not a line graph. It is a dot graph.

In junior high school mathematics we fail to mention the Cartesian product of the domain and range of the relation (or function), and, in doing so, we fail to impress the pupil with the idea that he is selecting a subset of the Cartesian product to define a relation (function). In most cases the Cartesian product consists of all the points in the plane. There are instances encountered in elementary work, however, in which the Cartesian product consists of all possible pairs of integers, or the Cartesian product of the set of integers and the set obtained by multiplying each integer by five. The latter Cartesian product would be encountered if we were to graph $C = 5n$ wherein n may be replaced by any positive integer. There is real reason to believe that graphing should be introduced with exercises in which the pupils are required to graph both the Cartesian product and the subset defining the relation. An example is: Graph all possible number pairs (a, b) which can be obtained from the sets $A = \{1, 2, 3, 4, 5\}$ and $B = \{5, 10, 15, 20, 25\}$; then find the subset for which $b = 5a$.

These examples and several given at the beginning of the chapter indicate the naturalness of using finite sets, finite Cartesian products, and graphs containing only a finite number of points in beginning algebra. Such sets are most appropriate for young children. Situations involving finite sets arise very often when we are dealing with objects that are not numbers. An example is the set of states and their populations given early in the chapter. There is no reason why we should not range the states along the x-axis and indicate their population by a graph. The Cartesian product idea enables us to broaden the graphing concept and to make it virtually independent of the pupils' algebraic skill.

ELEMENTARY FUNCTIONS IN SECONDARY SCHOOLS

The functions most commonly considered in the secondary schools (linear, quadratic, polynomial, trigonometric, exponential, and logarithmic) are widely used in science. The student should be thoroughly familiar with these functions. In this section we intend to suggest how the ideas of this chapter can be used to make learning more efficient

and give the student a deeper understanding of functions and relations through the study of these special functions.

Just as you study other things by looking at special situations—taking cross sections, so to speak—we study functions and learn about their special properties by placing restrictions on the set of number pairs (or on the first or second elements of the pairs). A simple example will illustrate how such restrictions may be helpful in the study of functions.

Consider the function defined by the rule '$y = 2x + 1$' ($x \in L$). Since $x \in L$, we know from all our properties of numbers that $y \in L$. So, our rule enables us to find number pairs in $L \times L$; that is, a set of real number pairs (x, y).

Suppose we select from this set those number pairs whose first element is zero; that is, the pairs $(0, y)$. To do this you write $y = 2 \cdot 0 + 1$; $y = 1$. This shows that the pair $(0, 1)$ is the only element in the function whose first member is 0. Graphically, we have found the number pair associated with the point at which the line crossed the y-axis. The y intercept is 1.

On the other hand, we may wish to select a subset of the domain, say $D = \{1, 2, 3, 4, 5, \cdots\}$ and study the corresponding subset R, of the range. Our rule tells us that to D there corresponds, element for element, $R = \{3, 5, 7, 9, 11, \cdots\}$. From this you *get the feel* that if you substitute for 'x' in '$y = 2x + 1$' (in succession) two numbers which differ by one, then the corresponding y-values will differ by two. Graphically, this is interpreted to mean that the slope of the line is 2. This slope is characteristic of the graph of '$y = 2x + 1$'. By suitable generalization we can bring the student to see that $\frac{y_1 - y_2}{x_1 - x_2}$ being constant is a property of any linear function.

The Linear Function. Consider the two-parameter family of functions defined by '$y = mx + b$', where x, m, and b are elements of L, and consequently $y \in L$. From this family select those pairs for which the first element is zero, that is, find all pairs $(0, y)$. Simple substitutions show that $(0, b)$ defines the pairs sought. In words, if we substitute '0' for 'x', we must substitute 'b' for 'y' to get a true statement. We call b the y intercept.

Next think of (x_1, y_1) and (x_2, y_2) as two number pairs which satisfy $y = mx + b$. Study the expression $\frac{y_1 - y_2}{x_1 - x_2}$, $(x_1 \neq x_2)$.

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{(mx_1 + b) - (mx_2 + b)}{x_1 - x_2} = \frac{m(x_1 - x_2)}{(x_1 - x_2)} = m.$$

That is, the ratio $\frac{y_1 - y_2}{x_1 - x_2}$ is always equal to m, regardless of the choice of (x_1, y_1) and (x_2, y_2). We say $\frac{y_1 - y_2}{x_1 - x_2}$ is constant for any linear function. Since this ratio is called the slope of the line, we have shown that the slope of the line associated with $y = mx + b$ is m and that the line crosses the y-axis at $(0, b)$.

The Quadratic Function. A quadratic function is defined by $y = ax^2 + bx + c$ where a, b, and c are elements of L and $a \neq 0$. Or, we may say it is the set of number pairs $\{(x, ax^2 + bx + c)\}$.

Which elements of the function have second members zero? You can answer this question by writing $0 = ax^2 + bx + c$. From this you see that

$$\left(\frac{-b + \sqrt{b^2 - 4ac}}{2a}, 0\right) \quad \text{and} \quad \left(\frac{-b - \sqrt{-4ac + b^2}}{2a}, 0\right)$$

are the number pairs sought provided $b^2 - 4ac > 0$. This work may be viewed in terms of inverses as suggested on pages 89 to 92. $(0, c)$ is the only element satisfying the condition $y = ax^2 + bx + c$ with first member 0. (Here c is the y intercept.)

Recall the role played by $\frac{y_1 - y_2}{x_1 - x_2}$ in the study of the linear function.

Using a similar attack on the quadratic function reveals that

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{a(x_1^2 - x_2^2) + b(x_1 - x_2)}{x_1 - x_2} = a(x_1 + x_2) + b$$

From Figure 28 you see that $\frac{y_1 - y_2}{x_1 - x_2}$ is the slope of a secant through $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$. Hence, from $\frac{y_1 - y_2}{x_1 - x_2} = a(x_1 + x_2) + b$ we

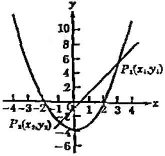


FIG. 28

have a formula for finding the slope of a secant through any two points of the parabola.

By taking the limit of $\frac{y_1 - y_2}{x_1 - x_2}$ as x_2 approaches x_1 we find the slope of the tangent at $P_1(x_1, y_1)$, that is,

$$\lim_{x_2 \to x_1} \frac{y_1 - y_2}{x_1 - x_2} = \lim_{x_2 \to x_1} [a(x_1 + x_2) + b] = 2ax_1 + b.$$

Hence, the slope of the tangent to a parabola at $P_1(x_1, y_1)$ is $2ax_1 + b$.

Setting $2ax_1 + b = 0$, we find $x_1 = -b/2a$. Since this is the value of 'x' where the tangent line has zero slope and is therefore horizontal, it is the x-coordinate of the maximum or minimum point of the parabola.

The Polynomial Functions. The general polynomial function is defined by $a_nx^n + a_{n-1}x^{n-1} + \cdots a_1x^1 + a_0 = y$ where n is a positive integer and the a_i are rational. Of course, the linear function and the quadratic function, previously discussed, are special cases of the general polynomial function.

The general polynomial is too complicated to study in a manner similar to the linear function and the quadratic function. However, a few special cases should be studied. Among these are:

(1) $y = ax^n$; n an even positive integer and $a > 0$. These curves have the general shape shown in Figure 29.

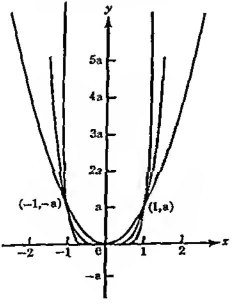


FIG. 29

If $a < 0$ these curves open downward with the vertex still at the origin.

(2) $y = ax^n$; n an odd positive integer and $a > 0$. This one parameter family of curves has the general shape shown in Figure 30.

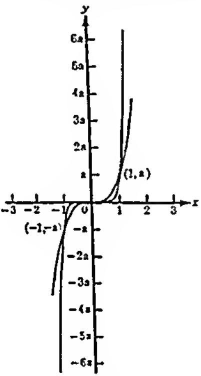


FIG. 30

If $a < 0$, these curves are rotated about the x-axis through 180°.

In the secondary schools the simplified polynomial functions defined by $y = ax^n$, where n may be replaced by any integer, are encountered under the heading of direct and inverse variation.

If numbers are paired according to the rule $y = ax^n$ for $n > 0$, then we say that y *varies directly as* x^n and, in particular, if $y = ax$, then y *varies directly as* x. Of course, this form may be generalized as illustrated in the following: (1) $y^2 = ax$. In this instance, y^2 varies directly as x, (2) $y^3 = ax^4$. Here y^3 varies directly as x^4, and (3) $y - b = a(x + c)$. Here $y - b$ varies directly as $x + c$.

The rule '$y = a/x^n$' leads to the terminology *y varies inversely as* x^n.

It is always instructive to have beginning classes form sets of number pairs via two rules such as '$y = 5x$' and '$y = x + 5$' and examine the

difference both graphically and *tablewise*. The one set of numbers leads to the direct variation terminology and the other does not. Graphing the two sets of pairs makes this difference very clear.

The Trigonometric Functions. While it is frequently the case that the trigonometric functions are introduced for the purpose of enabling the pupil to solve right triangles, you should keep in mind that the triangle solving property of trigonometric functions is relatively unimportant. It is vastly more important for the high school student to know that these functions are periodic. Their periodicity lies at the heart of the most important uses of these functions. Furthermore, they are the only functions studied in high school which are periodic.

Rather than discuss all six trigonometric functions we will use the sine function as an illustration of some of the important things to know about all trigonometric functions. We assume that the function defined by '$y = \sin x$', ($x \epsilon L$ and $y \epsilon L$) is familiar.

If we write '$y = a \sin bx$', we see that it defines a two-parameter family. In high school it is desirable to plot successive members of this family by assigning a value to 'b' and then assigning successive values to 'a'. You then obtain a family of curves as in Figure 31.

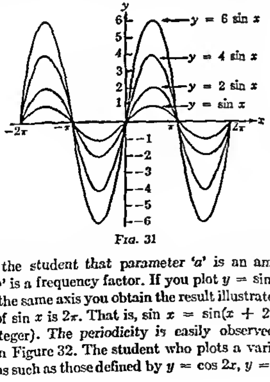


FIG. 31

This shows the student that parameter 'a' is an amplitude factor.

Parameter 'b' is a frequency factor. If you plot $y = \sin x$; $y = \sin 2x$; $y = \sin 3x$; on the same axis you obtain the result illustrated in Figure 32.

The period of $\sin x$ is 2π. That is, $\sin x = \sin(x + 2\pi) = \sin(x + 2n\pi)$ (n an integer). The *periodicity* is easily observed if you look at the graphs in Figure 32. The student who plots a variety of trigonometric functions such as those defined by $y = \cos 2x$, $y = \sin(x + \pi/2)$,

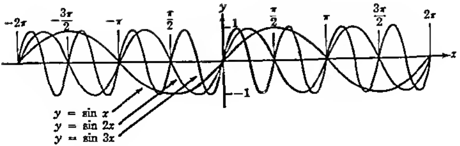


FIG. 32

$$y = 2 \sin x + 3 \sin 2x,$$

and $y = 3 \sin x + 2 \cos x$, will learn a great deal that will be very useful to him later in mathematics, in engineering, and in science. Students enjoy making these graphs, and tbe teacher can use such exercises to give practice in the use of tables, numerical computations, and trigonometric identities.

The solution of $y = \sin x$ for 'x' is not possihle if we restrict ourselves to the simple operations of addition, subtraction, multiplication, division, extraction of roots, and raising to powers. So, we create a new symbolism and a new function. We say that '$x = \sin y$' defines a new relation called the 'arc sine'. It is, of course, tbe inverse of the sine function, but it is not a function as is clear from Figure 33.

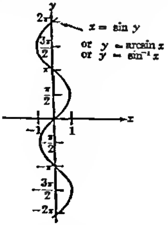


FIG. 33

For each value of 'x' in the range of this inverse relation (that is, $-1 \leq x \leq 1$), there is an infinite number of corresponding y-values. But we can consider this relation as the union of functions. In Figure 34

40 13

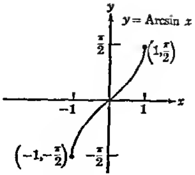


FIG. 34

we show one of these. It is called the principal value of the arc sine and is designated by writing the name of the inverse with a capital letter. Hence $y = \text{Arc sin } x$ defines a function that is part of the inverse of the sine function. All other values of y that satisfy $x = \sin y$ are given by $y = \text{Arc sin } x \pm 2k\pi$ or $y \pm 2k\pi = \text{Arc sin } x$, where k is any positive integer. We see from the graph that $-\pi/2 \leq \text{Arc sin } x \leq \pi/2$. Special names are not adopted for other branches of the inverse sine.

Sometimes, $y = \text{arc sin } x$ is written in place of $x = \sin y$. However, 'arc sin x' is an ambiguous expression. That is, for any given value of 'x', it might represent any one of the infinite number of values of y that satisfy $x = \sin y$. It is for this reason that it is better to use the principal value and get other solutions by the formulas given above.

The Exponential and Logarithmic Functions. The importance of the exponential and logarithmic functions transcends their uses in calculation. For this reason it is very important that the serious student of mathematics become familiar with their properties. In studying the exponential function defined by $y = a^x$ $(a > 1)$, the device of considering only a subset of the domain will be found useful. Simple bases such as $a = 2$, and special values of 'x' such as 1, 0, 2, -2, -10, suffice to get a very good idea of the function. By considering the ordered pairs, (1, 2), (2, 4), (3, 8), (4, 16), and so on, the pupil may be led to discover for himself that multiplying the second components may be accomplished by adding the first. Thus if we take the first two number pairs (1, 2) and (2, 4), we can add the two first elements $(1 + 2)$; look among the other pairs for a pair having 3 as a first element. This pair is (3, 8). The second element of (3, 8) is 2×4. Hence by adding the first elements of two pairs we have found the product of the second elements of these pairs

Such experiments lead naturally to considering the inverse defined by $x = a^y$. As in the case of the sine function, the elementary opera-

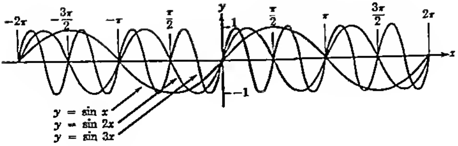


FIG. 32

$$y = 2 \sin x + 3 \sin 2x,$$

and $y = 3 \sin x + 2 \cos x$, will learn a great deal that will be very useful to him later in mathematics, in engineering, and in science. Students enjoy making these graphs, and the teacher can use such exercises to give practice in the use of tables, numerical computations, and trigonometric identities.

The solution of $y = \sin x$ for 'x' is not possible if we restrict ourselves to the simple operations of addition, subtraction, multiplication, division, extraction of roots, and raising to powers. So, we create a new symbolism and a new function. We say that '$x = \sin y$' defines a new relation called the 'arc sine'. It is, of course, the inverse of the sine function, but it is not a function as is clear from Figure 33.

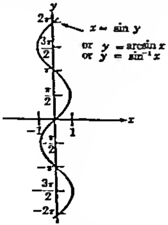


FIG. 33

For each value of 'x' in the range of this inverse relation (that is, $-1 \leq x \leq 1$), there is an infinite number of corresponding y-values. But we can consider this relation as the union of functions. In Figure 34

40 13

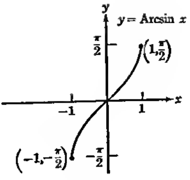


FIG. 34

we show one of these. It is called the principal value of the arc sine and is designated by writing the name of the inverse with a capital letter. Hence $y = \text{Arc sin } x$ defines a function that is part of the inverse of the sine function. All other values of y that satisfy $x = \sin y$ are given by $y = \text{Arc sin } x \pm 2k\pi$ or $y \pm 2k\pi = \text{Arc sin } x$, where k is any positive integer. We see from the graph that $-\pi/2 \leq \text{Arc sin } x \leq \pi/2$. Special names are not adopted for other branches of the inverse sine.

Sometimes, $y = \text{arc sin } x$ is written in place of $x = \sin y$. However, 'arc sin x' is an ambiguous expression. That is, for any given value of 'x', it might represent any one of the infinite number of values of y that satisfy $x = \sin y$. It is for this reason that it is better to use the principal value and get other solutions by the formulas given above.

The Exponential and Logarithmic Functions. The importance of the exponential and logarithmic functions transcends their uses in calculation. For this reason it is very important that the serious student of mathematics become familiar with their properties. In studying the exponential function defined by $y = a^x$ $(a > 1)$, the device of considering only a subset of the domain will be found useful. Simple bases such as $a = 2$, and special values of 'x' such as 1, 0, 2, −2, −10, suffice to get a very good idea of the function. By considering the ordered pairs, (1, 2), (2, 4), (3, 8), (4, 16), and so on, the pupil may be led to discover for himself that multiplying the second components may be accomplished by adding the first. Thus if we take the first two number pairs (1, 2) and (2, 4), we can add the two first elements $(1 + 2)$; look among the other pairs for a pair having 3 as a first element. This pair is (3, 8). The second element of (3, 8) is 2×4. Hence by adding the first elements of two pairs we have found the product of the second elements of these pairs

Such experiments lead naturally to considering the inverse defined by $x = a^y$. As in the case of the sine function, the elementary opera-

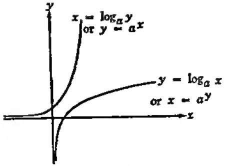


FIG. 35

tions do not suffice to solve for 'y'. But the inverse is a function, as indicated in Figure 35. Hence, we adopt a symbol '$y = \log x$' for the solution of $x = a^y$. The inverse of $y = a^x$ is defined by $y = \log_a x$ or by $x = a^y$.

By using simple bases and restricted domains these two functions could be studied by very young pupils. After such rich experiences, they would find the table of logarithms most natural. For after all, a table of common logarithms is merely a more detailed listing of certain pairs of numbers whose second members are the logarithms of the first members; i.e., $(x, \log x)$. The first components are listed in the margins and the second components in the body of the table. By reversing the point of view and considering the first components to be in the body of the table, we have a table of powers of 10. This point of view could do much to remove the mystery from characteristics and mantissas and to produce students able to use logarithms in accordance with common sense.

Other Important Functions. The elementary functions discussed above have a time-honored and well-justified position in the secondary curriculum. But there are many other interesting functions with worthwhile applications that are suitable for treatment in the elementary and secondary schools. We have given some examples in earlier parts of the chapter by using finite domains, and the reader can construct any number of examples.

Too often students get the idea that a function is always defined by a simple formula giving 'y' in terms of an elementary formula involving 'x'. This is not so as we have seen in numerous examples in this chapter. Any set of ordered pairs is a relation. There are always many rules defining the relation, and it may be that there is no rule involving only elementary functions. The function concept should be presented, and used, in its complete generality in the high school. We mention here a few examples of important functions that will help widen the student's horizon.

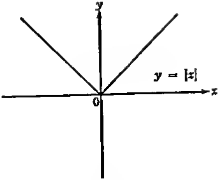


FIG. 36

In Figure 36 is shown the graph of $y = |x|$. For $x \geq 0$, the graph coincides with the graph of $y = x$. For $x \leq 0$, the graph coincides with the graph of $y = -x$. Familiarity with this graph (and with many others that can be drawn once it is understood; for example, $y = |x| + 1$ and $y = |x| + x$) will help the student master the important concept of absolute value.

There is no reason why the rule for ordering pairs should be the same rule throughout the domain of definition of the function. For example: $y = x$ for $0 < x < 4$, and $y = 4$ for $4 \leq x$ defines a perfectly good function. This graph is as shown in Figure 37.

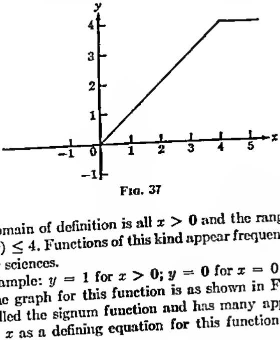


FIG. 37

Here the domain of definition is all $x > 0$ and the range of the function is $0 < f(x) \leq 4$. Functions of this kind appear frequently in psychology and other sciences.

Another example: $y = 1$ for $x > 0$; $y = 0$ for $x = 0$; and $y = -1$ for $x < 0$. The graph for this function is as shown in Figure 38. This function is called the signum function and has many applications. We write $y = \text{sg}\ x$ as a defining equation for this function.

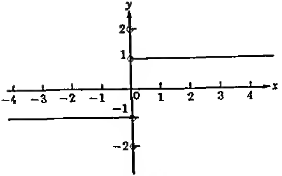


Fig. 38

The function graphed in Figure 6 (page 76) is called a step function. Many step functions can be defined in terms of the notation, '$[x]$', which means the greatest integer less than or equal to x. The function of Figure 6 is given by $C = 3[w] + 3$, $w \neq 1, 2, 3, \cdots$, or $C = 3w$, $w = 1, 2, 3, \cdots$.

An enormous variety of functions can be defined by using the above functions or any rule suggested by scientific data and the level of the student. Such functions may be utilized to give drill in algebra, graphing, and trigonometry in contexts that intrigue the student. For example: after graphing $y = \sin x$, try $y = \text{sg}(\sin x)$.

SOME GENERALIZATIONS OF THE DEFINITION OF FUNCTION

Binary Operations. In the last few sections of this chapter, we have been considering important classes of functions which should be studied in a high school program. There are other classes of functions which should be mentioned briefly even though they play little or no part in the present day secondary program.

In mathematics the term 'binary operation' occurs frequently. Elementary school pupils are familiar with binary operations although they do not know the term. The basic addition facts describe the binary operation $+$ on the set of integers. The operation is binary because only two integers are involved.

Now it is possible to show that binary operations are special instances of the function idea; i.e., they can be defined in terms of sets of ordered pairs where the first element of the pair is itself a pair of numbers. An example will help to clarify this idea. Let I represent the set of positive integers. Form the set $A = \{[(x, y), z] \mid x \in I, y \in I, z \in I, \text{and } x + y = z\}$. Some elements of this set are: $[(1, 1), 2]$, $[(1, 2), 3]$, $[(4, 6), 10]$. Stated in the old familiar symbolism we have: $1 + 1 = 2$, $1 + 2 = 3$, and

$4 + 6 = 10$. So, we see that it is possible to view A as a function with pairs of numbers as first elements and single numbers as second elements. A is the collection of all the addition facts in arithmetic.

With this illustration it will be easier to comprehend the more precise definition; namely, a binary operation on a set K is a function whose domain is $K \times K$ and whose range is K or a subset of K. We usually speak of a binary operation from $K \times K$ to K. To show how this definition works, take K as the set of all integers and zero. Hence, $K \times K$ is the set of all pairs of integers (zero included in the pairings). Now associate each of these pairs (x, y) with an element of $K(z)$ according to the rule $x + y = z$. We now have a set $\{[(x, y), z]\}$ which defines the binary operation $+$ from $K \times K$ to K. Note that for each first element (x, y) there is one, and only one, second element z. Hence, the binary operation is a function. It is easy to construct sets $\{[(x, y), z]\}$ according to the rule $xy = z$ or $x + 2y = z$. These sets are functions. On the other hand the set $\{[(x, y), z]\}$ constructed according to the rule '$z^2 = x + y$' is a nonfunctional relation—a ternary relation.

It is easy to generalize these ideas on binary relations, ternary relations, binary functions, and so on, to ordered n-tuples but this will be left as an exercise (or further reading) for those who are interested.

Set Functions. In many instances it is desirable to assign numbers to sets. Thus, if $A = \{1, 3, 10, 100\}$ we can assign the number 4 to A; i.e., the number of elements in A. Let us write $|A| = 4$. Of course there are many rules we can use to assign numbers to sets. The average grade in a class; the height of the tallest boy in a class; the largest bank in a state; and many other rules that easily occur to anyone can be used to assign a number to a set of things.

Consider $A = \{0, 1, 2, 3\}$. Let us write all the possible subsets of A. They are: $A_1 = \{0\}$, $A_2 = \{1\}$, $A_3 = \{2\}$, $A_4 = \{3\}$, $A_5 = \{0, 1\}$, $A_6 = \{0, 2\}$, $A_7 = \{0, 3\}$, $A_8 = \{1, 2\}$, $A_9 = \{1, 3\}$, $A_{10} = \{2, 3\}$, $A_{11} = \{0, 1, 2\}$, $A_{12} = \{0, 1, 3\}$, $A_{13} = \{0, 2, 3\}$, $A_{14} = \{1, 2, 3\}$,

$$A_{15} = A = \{0, 1, 2, 3\}$$

and $A_{16} = \{\ \}$. Note that there are 2^4 subsets of A and that we call A a subset of A as well as A_{16}, the null set. This arrangement provides a universe of discourse, $\{A_i\}$, and if we assign to each subset of A the number corresponding to the number of elements in the subset we then have a relation—a set relation. It is $\{(A_i, N_i) \mid A_i \subset A \text{ and } |A_i| = N_i\}$. A few elements of this relation (function) are $(A_1, 1)$, $(A_2, 1)$, $(A_7, 2)$, $(A_{16}, 0)$. In words, a set function is a set of ordered pairs whose first elements are subsets of a given set A and whose second elements are unique and assigned to these subsets according to some rule.

Possibly the widest use of set functions occurs in the study of probability. Consider a collection of four balls in a bag. Also, assume there is just one ball of each of these colors: white, green, blue, and yellow. Designate this collection by $A = \{w, g, b, y\}$. Let us draw balls from the bag and call each such possible subset of A an event. Thus drawing the subset $\{w, g\}$ will be an event. There are sixteen possible events in this instance. Also, assume that it is equally likely that each ball will be drawn and that after each draw the ball is replaced and the bag shaken. The probability of the event $\{w\}$ occurring is obviously $\frac{1}{4}$. We will write $P\{w\} = \frac{1}{4}$, $P\{w, g\} = \frac{2}{4}$ (the probability of drawing either a white or a green ball); $P\{w, g, b, y\} = 1$; and $P\{\ \} = 0$.

So we see that for each subset of A we have assigned a number p such that $0 \leq p \leq 1$ and we call this set of pairs (subsets of A and probabilities) a set function and in particular, a probability measure. It is easy to set up other set functions. This you may wish to do.

SUMMARY

Webster's New International Dictionary defines 'function' in this way: "A magnitude so related to another magnitude that to values of the latter there correspond values of the former." The words 'magnitude' and 'correspond' are key words in this definition. What do they mean? Whatever the answer to this question, the notion is vague. This being so, the definition cannot be used by a mathematician. It does not satisfy the requirements for precise statements demanded by the mathematical world. Neither does such a statement satisfy the requirements of good teaching. Vague statements do not facilitate communication between pupil and teacher.

In contrast, the definition based on set considerations is precise and clear. A function or a relation is a set of ordered pairs. This is a definite entity; one you can almost put your hands on. This being the case, it would seem logical that it be considered as the basis for instruction in elementary mathematics. Those teachers with venturesome spirits may wish to introduce the set definitions for relation and function in one or more of their classes. The results may be rewarding.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in this chapter.

4

Proof

EUGENE P. SMITH AND KENNETH B. HENDERSON

SOME COMMON USES OF THE TERM *PROOF*

"PROOF of the superior quality of moron gasoline! Three billion miles have been driven on Oshkosh's new *moron* gasoline." Do three billion miles of driving prove the superior quality of the gasoline as proclaimed by this advertisement? Or, how about this one. "Proof positive. Out of three hundred people tested, 4 out of 5 preferred Llsma cigarettes for mildness, coolness, and plain good smoking." Just what constitutes proof anyway?

One boy says to another, "I'll prove which one is the best fighter. Come on out in the alley, I'll show ya!" He does win the fisticuffs. Is this proof?

Or, let us change the scene to that of a courtroom. For days the prosecuting attorney sends witness after witness to the stand to fix the finger of guilt on the defendant charged with murder. The defense attorney cross-examines the witnesses and sends his own parade of defense witnesses to the stand to try to establish the innocence of his client. The jury listens to all of the evidence presented, and then at the end of the trial retires to another room to decide upon a verdict. Will the jurors likely be dealing with proof or with probability?

How do these illustrations compare with that of the second grade youngster who is in an argument over the answer to the subtraction problem "27 − 12"?

"See, John, I'm right, because fifteen plus twelve equals twenty-seven."

This child is using a definition of subtraction (intuitively, of course) to establish the validity of his answer. Assuming no mistakes in his addition, does his argument represent proof or simply probability?

Proof is a common word in our vocabularies with various shades of meaning in its daily usage, but it has a very special and precise meaning in mathematics. As a mature concept, proof in mathematics is a sequence of related statements directed toward establishing the validity of a

Possibly the widest use of set functions occurs in the study of probability. Consider a collection of four balls in a bag. Also, assume there is just one ball of each of these colors: white, green, blue, and yellow. Designate this collection by $A = \{w, g, b, y\}$. Let us draw balls from the bag and call each such possible subset of A an event. Thus drawing the subset $\{w, g\}$ will be an event. There are sixteen possible events in this instance. Also, assume that it is equally likely that each ball will be drawn and that after each draw the ball is replaced and the bag shaken. The probability of the event $\{w\}$ occurring is obviously $\frac{1}{4}$. We will write $P\{w\} = \frac{1}{4}$, $P\{w, g\} = \frac{2}{4}$ (the probability of drawing either a white or a green ball); $P\{w, g, b, y\} = 1$; and $P\{\ \} = 0$.

So we see that for each subset of A we have assigned a number p such that $0 \leq p \leq 1$ and we call this set of pairs (subsets of A and probabilities) a set function and in particular, a probability measure. It is easy to set up other set functions. This you may wish to do.

SUMMARY

Webster's New International Dictionary defines 'function' in this way: "A magnitude so related to another magnitude that to values of the latter there correspond values of the former." The words 'magnitude' and 'correspond' are key words in this definition. What do they mean? Whatever the answer to this question, the notion is vague. This being so, the definition cannot be used by a mathematician. It does not satisfy the requirements for precise statements demanded by the mathematical world. Neither does such a statement satisfy the requirements of good teaching. Vague statements do not facilitate communication between pupil and teacher.

In contrast, the definition based on set considerations is precise and clear. A function or a relation is a set of ordered pairs. This is a definite entity; one you can almost put your hands on. This being the case, it would seem logical that it be considered as the basis for instruction in elementary mathematics. Those teachers with venturesome spirits may wish to introduce the set definitions for relation and function in one or more of their classes. The results may be rewarding.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in this chapter.

4

Proof

EUGENE P. SMITH AND KENNETH B. HENDERSON

SOME COMMON USES OF THE TERM *PROOF*

"PROOF of the superior quality of moron gasoline! Three billion miles have been driven on Oshkosh's new *moron* gasoline." Do three billion miles of driving prove the superior quality of the gasoline as proclaimed by this advertisement? Or, how about this one. "Proof positive. Out of three hundred people tested, 4 out of 5 preferred Llama cigarettes for mildness, coolness, and plain good smoking." Just what constitutes proof anyway?

One boy says to another, "I'll prove which one is the best fighter. Come on out in the alley, I'll show ya!" He does win the fisticuffs. Is this proof?

Or, let us change the scene to that of a courtroom. For days the prosecuting attorney sends witness after witness to the stand to fix the finger of guilt on the defendant charged with murder. The defense attorney cross-examines the witnesses and sends his own parade of defense witnesses to the stand to try to establish the innocence of his client. The jury listens to all of the evidence presented, and then at the end of the trial retires to another room to decide upon a verdict. Will the jurors likely be dealing with proof or with probability?

How do these illustrations compare with that of the second grade youngster who is in an argument over the answer to the subtraction problem "27 − 12"?

"See, John, I'm right, because fifteen plus twelve equals twenty-seven."

This child is using a definition of subtraction (intuitively, of course) to establish the validity of his answer. Assuming no mistakes in his addition, does his argument represent proof or simply probability?

Proof is a common word in our vocabularies with various shades of meaning in its daily usage, but it has a very special and precise meaning in mathematics. As a mature concept, proof in mathematics is a sequence of related statements directed toward establishing the validity of a

conclusion. Each conclusion is (or can be) justified by reference to recognized and accepted assumptions (including the assumptions of logic), definitions, undefined terms, previously proved propositions (including those of logic), or a combination of these reasons. None of the above illustrations is an example of proof in the mathematical sense except the situation involving the second grader. The little boy's conclusion rests on a definition of subtraction; i.e., in general $a - b = c$ if and only if $a = b + c$. Since the condition $a = b + c$ ($27 = 12 + 15$) is satisfied, the answer has been *proved* to be correct. Thus we have a very simple case of proof as the mathematician views it. The conclusion in each of the other instances is dependent upon what is commonly called induction. In each of these latter cases there is a necessity to collect empirical data concerning the situation, and the ultimate decision in each case rests upon the accumulated evidence. Mathematical proof, on the other hand, is independent of observation and experimentation except as these processes may suggest assumptions which are accepted by the mathematician as unproved propositions, and as they suggest conjectures to be proved.

Proof, broadly conceived in the teaching of elementary and secondary school mathematics, is an ever-expanding concept growing from the immature preschool and elementary school stage of *that which convinces* to the mature concept of proof as defined above which Keyser so aptly calls *if-then* or *postulational thinking*.[1] As Whitehead states in his stimulating lectures on *The Function of Reason*, "We all start by being empiricists."[2] The genesis of the mature concept of proof lies deep in the empiricism, that is, in experimentation and observation of early childhood.

In the ensuing discussion the processes of induction and deduction occupy the spotlight. The importance of thorough teaching of these concepts was emphasized in the Final Report of the Joint Commission in the following impressive manner.[3]

In secondary instruction there should be conscious experience with both inductive and deductive reasoning. The character and requisites of these two procedures should be so clearly grasped that appropriate behavior on the part of the pupil is brought about in the direction both of understanding and of making applications. In solving problems the pupil should develop the habit of asking whether he is starting from general premises and is seeking consequences, or, by examining particular instances, is aiming at universal conclusions. He should seek to discover and remove ambiguity in the use of terms. He should understand the relation between assumptions and conclusions, and he should grow in the ability to judge the validity of reasoning which purports to establish proof. Proper attention must be given to generalizing these behaviors and

understandings. We may then hope that pupils will apply them to situations arising in many different fields of thought.

The rest of this chapter will be devoted to an analysis of the nature of probable inference (induction) and necessary inference (deduction), a presentation of some techniques used in each, and the relationships and differences between the two types of inference. Illustrations have been selected to demonstrate that mathematics can be taught in ways that will help children learn to test the implications of ideas, develop accepted criteria for the adequacy of a proposed proof, and to gain some facility in proving relationships themselves.

In general, the first part of each of two major sections of this chapter, *Probable Inference* and *Necessary Inference*, is devoted to ideas that may be taught in the elementary grades. The material in each section becomes progressively more advanced in terms of grade level and in terms of the concept of proof. The reader who reaches the limit of his interest in the first part of the chapter should turn to page 140 where the discussion of elementary ideas of necessary inference begins.

PROBABLE OR INDUCTIVE INFERENCE

Before examining probable or inductive inference in some detail, let us note a few generalizations and raise some questions about the nature of probable inference. The premises for conclusions reached by probable inference are factual data or evidence which are collected through the five senses.

You no doubt recognize that induction is at the very heart of the scientific method, and hence to its procedures must go much of the credit for the tremendous advance made by civilization in this scientific age. However, one needs only to look briefly at the history of science to realize that many generalizations arising out of the organization and analysis of sense data are only tentative. They must be dated; that is, they represent our conclusions as of some given time, say January 1, 1959. New evidence or new insights may result in a changed conclusion.

If this is so, how can one be sure that a given proposition reached through induction is true? How can one be certain that predictions based on the proposition will agree with the observed facts? This is one of the very difficult questions of probable inference; it is the problem of the verification of a conjecture, guess, or hypothesis (as the word is used in science). We shall have more to say about verification later in this chapter.

With this brief introduction we shall plunge directly into a discussion of some elementary ideas of probable inference. The topic of probability

conclusion. Each conclusion is (or can be) justified by reference to recognized and accepted assumptions (including the assumptions of logic), definitions, undefined terms, previously proved propositions (including those of logic), or a combination of these reasons. None of the above illustrations is an example of proof in the mathematical sense except the situation involving the second grader. The little boy's conclusion rests on a definition of subtraction; i.e., in general $a - b = c$ if and only if $a = b + c$. Since the condition $a = b + c$ $(27 = 12 + 15)$ is satisfied, the answer has been *proved* to be correct. Thus we have a very simple case of proof as the mathematician views it. The conclusion in each of the other instances is dependent upon what is commonly called induction. In each of these latter cases there is a necessity to collect empirical data concerning the situation, and the ultimate decision in each case rests upon the accumulated evidence. Mathematical proof, on the other hand, is independent of observation and experimentation except as these processes may suggest assumptions which are accepted by the mathematician as unproved propositions, and as they suggest conjectures to be proved.

Proof, broadly conceived in the teaching of elementary and secondary school mathematics, is an ever-expanding concept growing from the immature preschool and elementary school stage of *that which convinces* to the mature concept of proof as defined above which Keyser so aptly calls *if-then* or *postulational thinking*.[1] As Whitehead states in his stimulating lectures on *The Function of Reason*, "We all start by being empiricists."[2] The genesis of the mature concept of proof lies deep in the empiricism, that is, in experimentation and observation of early childhood.

In the ensuing discussion the processes of induction and deduction occupy the spotlight. The importance of thorough teaching of these concepts was emphasized in the Final Report of the Joint Commission in the following impressive manner.[3]

> In secondary instruction there should be conscious experience with both inductive and deductive reasoning. The character and requisites of these two procedures should be so clearly grasped that appropriate behavior on the part of the pupil is brought about in the direction both of understanding and of making applications. In solving problems the pupil should develop the habit of asking whether he is starting from general premises and is seeking consequences, or, by examining particular instances, is aiming at universal conclusions. He should seek to discover and remove ambiguity in the use of terms. He should understand the relation between assumptions and conclusions, and he should grow in the ability to judge the validity of reasoning which purports to establish proof. Proper attention must be given to generalizing these behaviors and

understandings. We may then hope that pupils will apply them to situations arising in many different fields of thought.

The rest of this chapter will be devoted to an analysis of the nature of probable inference (induction) and necessary inference (deduction), a presentation of some techniques used in each, and the relationships and differences between the two types of inference. Illustrations have been selected to demonstrate that mathematics can be taught in ways that will help children learn to test the implications of ideas, develop accepted criteria for the adequacy of a proposed proof, and to gain some facility in proving relationships themselves.

In general, the first part of each of two major sections of this chapter, *Probable Inference* and *Necessary Inference*, is devoted to ideas that may be taught in the elementary grades. The material in each section becomes progressively more advanced in terms of grade level and in terms of the concept of proof. The reader who reaches the limit of his interest in the first part of the chapter should turn to page 140 where the discussion of elementary ideas of necessary inference begins.

PROBABLE OR INDUCTIVE INFERENCE

Before examining probable or inductive inference in some detail, let us note a few generalizations and raise some questions about the nature of probable inference. The premises for conclusions reached by probable inference are factual data or evidence which are collected through the five senses.

You no doubt recognize that induction is at the very heart of the scientific method, and hence to its procedures must go much of the credit for the tremendous advance made by civilization in this scientific age. However, one needs only to look briefly at the history of science to realize that many generalizations arising out of the organization and analysis of sense data are only tentative. They must be dated; that is, they represent our conclusions as of some given time, say January 1, 1959. New evidence or new insights may result in a changed conclusion.

If this is so, how can one be sure that a given proposition reached through induction is true? How can one be certain that predictions based on the proposition will agree with the observed facts? This is one of the very difficult questions of probable inference; it is the problem of the verification of a conjecture, guess, or hypothesis (as the word is used in science). We shall have more to say about verification later in this chapter.

With this brief introduction we shall plunge directly into a discussion of some elementary ideas of probable inference. The topic of probability

as it is treated numerically and with logical rigor in mathematics is a tricky and difficult one. In the past very little of this has been taught in secondary schools.

Modern developments in mathematics and its uses are leading many people to believe that more should be taught, but exactly what should be included and how it should be presented is still quite uncertain. Chapter 6, on probability, contains extensive suggestions for the teaching of this topic at elementary and secondary school levels. In the inductive reasoning or probable inference part of this chapter the emphasis will be placed on developing an intuitive sense or feeling for the distinction between that which is probably true and that which has been proven. This may provide students with both a foundation for a later more numerical and precise treatment of probability and a better understanding of the nature of proof. This may also suggest to teachers devices for leading students to *discover* facts and relationship. Deductive inference or proof is then the device which converts such discoveries from conjectures which are probably true to established theorems. The importance of induction to mathematicians as a source of fruitful conjectures is often obscured by their emphasis on logic. There are several typical sources of probable inferences.

Recognition of Authority

At school, children often close, or attempt to close, an argument with the statement, "My Dad (Mother) says so, that's why!" At home, the tables are frequently turned, for the child considers it sufficient in arguing a point with a playmate or his parents to counter with, "I know it is true, because my teacher said so." These statements reveal one of the first sources of convincing evidence to a child—the recognition of authority. A youngster is also likely to recognize as authorities any adults whom he knows and trusts (these may be personalities whom he *knows* via radio and television as well as personal acquaintances), and, after he learns to read, he tends to accept as true any material that he reads that he does not label as "*just a story*." Gesell and Ilg point out that up to eight years of age children even have difficulty differentiating between fantasy and reality. This is also evidenced by the fact that they must be reminded frequently when they are listening to a story that it is just *make believe* or that it is *real*. This confusion of fantasy and reality complicates any choice these youngsters must make among authorities. Gradually the child begins to recognize conflicts among authorities, to question intelligently the generalizations which he hears or reads, and at about nine years of age independent critical thinking emerges.[4]

It should be noted, however, that it is much easier to criticize the statements of others than it is to be critical and objective about one's own statements. The child at this level still does not generalize from his own experience on the same level with which he can criticize the judgments of others. In the elementary school, particularly, children need considerable direction and assistance in projects and experiments, in organizing data, recognizing the relationships involved, and expressing these relations in language that they can clearly understand. Without this guidance, these children have little chance to think for themselves and must remain on a low level of blind dependence upon authority as the sole source of convincing evidence.

The most important authorities to a child who is learning mathematics are his teacher of arithmetic in the elementary school, his teachers of mathematics in the secondary school, and the authors of textbooks used by the pupil. In some cases the parents play important roles in this respect as they help their offspring with homework and to capitalize upon out-of-school experiences which promote the learning of quantitative and spatial concepts. But the child himself as he grows to maturity must become competent in judging the validity of information and inferences, for in our highly mathematized culture we are continuously beset with statistics, statistical generalizations, and subjective judgments in support of propositions that we are pressed to accept.

Seeing Is Believing

Children are firm adherents to the old adage that *seeing is believing.* "Well, I saw it with my own eyes, I ought to know," is a favorite justification of children for the validity of evidence. The things they *see* may range from ghosts to space ships, but they always seem to speak with conviction. The use of visual evidence is an important stage of development, for it is through this medium of seeing and handling objects that the child first learns his arithmetic. He knows that three plus four equals seven because when he combines a group of three objects with a group of four objects, by actual count there are seven objects in the new group. He can *see* this. In working with common fractions he can learn by visual means using commercial, teacher-made, or pupil-made devices that one fourth added to one fourth is equal to one half. He can see that three fourths minus one twelfth is two thirds. These visual aids can be used to verify the results of using the common algorithms concerning operations with fractions. Better yet, they can be used to help children establish the algorithms for themselves. Piaget emphasized the importance of such experiences when he observed that "From 7–8 to 8–12 years 'concrete operations' are organized, i.e., opera-

tional groupings of thought concerning objects that can be manipulated or known through the senses."[5] He goes further to emphasize the importance of visual schemas in children's thought as a step toward more abstract thought that is independent of visual schemas of concrete objects.

A variation of the *seeing is believing* approach may depend on a rearrangement of materials or ingenuity in adding the right lines to a figure. That this may be true on varying levels of maturity is illustrated by the following problems from informal geometry. A pupil wishes to find the area of a rectangle, the dimensions of which are expressed in an integral number of inches. He might place an inch-square on the rectangle in a systematic fashion until he had determined, by keeping a tally, how many inch-squares could have been placed on the rectangle. Thus he would have determined its area. (This plan involves measurement and counting concepts also.) Or he might have used the more efficient system of drawing a family of lines one inch apart and parallel to one side and another family of lines one inch apart and parallel to one of the sides perpendicular to the first one. This would have divided the rectangle into inch-squares. Again he could count these inch-squares to determine the number of square inches in the rectangle. By measuring the lengths and widths of a few rectangles, dividing them into inch-squares (or fractions thereof if the lengths of the sides involve fractional parts of the units used) and counting to determine the total number of square units in each rectangle, any sixth or seventh grader can derive the familiar relationship that the number of units of square measure in the area of a rectangle is equal to the product of the numbers of linear units in its length and its width. (For the sake of brevity this statement is often shortened to *The area of a rectangle is equal to the product of the length and the width of the rectangle.*) The pupils feel secure in their process of reasoning and have no doubt of the truth of their generalization which rests upon the empirical data that they themselves have collected.

It is common practice to assume the formula for the area of a rectangle in demonstrative geometry because a proof involves the theory of limits. This assumption is then used as a basis for proving the formulas for finding the areas of the other rectilinear figures. If this is done, the students' attention should be called to the nature and importance of this unproved assumption and to the fact that their earlier experimentation was not a proof but a verification that such a formula would be an appropriate one.

To illustrate how a rearrangement of materials may lead to further generalizations, consider now the problem of the pupil who was trying to

get a formula for the area of a parallelogram. Attempts to divide the parallelogram into inch-squares met with little success because the sides of the original figure were not perpendicular and there were too many three- and five-sided figures formed by the sides of the parallelogram which required estimation to determine what fraction of a square inch they were. He then reasoned: "I know how to find the area of a rectangle. Could there be a relation between the area of the parallelogram and the area of a rectangle?" After some thought and experimentation, it suddenly dawned upon him that he could cut a right triangle off one end of the parallelogram, put it on the other end, and *it appears* to form a rectangle (Fig. 1). He noted that the length of the newly formed rectangle

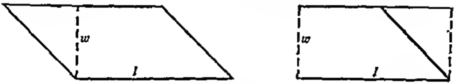


FIG. 1

was equal to that of the base of the parallelogram, but its width was the perpendicular distance between the base of the parallelogram and the side parallel to it, *not* the length of a side at all! He then proudly pronounced, "The area of a parallelogram is equal to the product of the lengths of its base and its width." This *width*, of course, is what we would call the height or altitude of the parallelogram. (A more precise and correct statement of the conclusion might be as follows: "The number of units of square measure in the area of a parallelogram is equal to the product of the numbers of units of linear measure in its base and in its altitude.") A similar procedure can be used to develop the formulas for the areas of the triangle, trapezoid, and circle. It will be shown later that regardless of how convincing these demonstrations may have been to the child, this process of reaching a generalization still does not constitute proof in the mathematical sense of the word. At this maturity level, however, evidence gained by empirical means is sufficient to permit the child to generalize the formulas for finding the areas of common rectilinear figures and to convince him of the truth of these formulas. Further, he was using *if-then* reasoning in order to extend his empirically derived formula to additional figures.

In demonstrative geometry, fruitful conjectures may be suggested to students by appropriate drawings. In a triangle, accurate constructions suggest the conjectures that: (1) the angle bisectors are concurrent, (2) the medians are concurrent, (3) the altitudes are concurrent, and (4) the perpendicular bisectors of the sides are concurrent. There are also

many occasions when a carefully constructed diagram may be sufficient to discredit a conjecture which happens to be false. For example, geometry students often think that the bisector of an angle of any triangle will also bisect the side opposite the bisected angle. Bisecting the obtuse angle of a scalene obtuse triangle with one small acute angle and one comparatively large acute angle will be sufficient to convince the most ardent exponent of the conjecture that any attempts to prove his original statement would be a waste of time. The understanding that a single *counter example* is enough to disprove a conjecture is an important part of an understanding of the nature of proof.

Measurement

Measurement provides another avenue for children to arrive at *truth*. This procedure also involves the element of *seeing is believing* discussed above. For example, one elementary school class determined the relationship among pints, quarts, and gallons by using standard measures of these sizes and pouring water from one to the other. Data could readily be obtained at a more advanced level for finding the approximate relationships between metric units and their corresponding English units of measure. Rather than being told the conversion factors for changing one unit of measure into another, pupils are given the opportunity to discover the approximate relationships and to generalize them for themselves. Such processes are pedagogically desirable because using concrete materials and having pupils discover relationships increases their understanding and retention of the ideas. However, a careful discussion of the nature of units as established by arbitrary definitions is also important and makes this process a place where one can also begin to build an understanding of the nature and role of definitions in a logical system.

At a somewhat higher level and in a conceptually different situation, this method suffices for determining the formula for the volume of a pyramid when the formula for finding the volume of a prism is known. The student obtains or makes two containers, one a prism and the other a pyramid. The prism and the pyramid should have equal heights and bases of equal area. (It is more obvious to the experimenter if the bases are congruent.) The problem of finding how many times the volume of the pyramid is contained in the volume of the prism can then be solved by using some appropriate substance, such as water, salt, or sawdust, which can be poured from one container to the other. Children who do not already know the relationship will seldom guess that it will take three full pyramids to fill the corresponding prism. Such experimentation

with the consequent surprise ending makes a lasting impression upon children. This same method can be used to find the relationship between the volumes of a cone and a cylinder.

Many teachers of geometry use exercises which involve measurement of angles or of line segments to suggest conjectures for formal proofs. For example, homemade devices or commercially-made teaching aids may be used to suggest the conjectures relating an inscribed angle to its intercepted arc, or the angle between two secants to the same circle and the difference of the two arcs of a circle intercepted between the two secants. Tables of data are made of readings taken from the devices, and conjectures which then seem reasonable to the student are subjected to the strict test of rigorous proof.

Again it should be emphasized that regardless of the volume of supporting evidence collected by measurement, this does *not* prove the conjecture in the mature mathematical sense of the word *prove*.

The Pragmatic Method

A fourth grade youngster was making a turtle trap. Eight or ten holes had been bored along each side of a square wooden frame. When this job was finished the child discovered that the dowel rods which he had selected would not go through the holes he had made for them.

"I can't understand. I thought they would fit. I measured," the boy exclaimed.

When asked how he measured the boy replied, "By my eye. I was sure they were the same size."

In order to move ahead on the project the child selected a larger bit, bored one hole, then checked to see if one of the dowel rods would fit. Since it did, he proceeded to finish his project.

This anecdote illustrates what might be called a pragmatic or *does it work* type of reasoning so common in children's thinking. The algebra student who is using the *guess and check* method in his initial work in solving the equation, $3x - 4 = x + 6$, is using this pragmatic method. For example, he may guess that the x equals six. Substituting *six* for x in the original equation does not result in a true statement; therefore he knows his first guess is wrong. Repeating this procedure with a second hypothesis that x equals *five*, results in a true statement; the boy concludes then that x equals *five* is a solution to the equation.

Counting

"In its simplest form, induction may be illustrated by the simple process of counting, and the propositions announcing the result would

be merely descriptive."[8] Not only is it the simplest, but counting is the most frequently used of all inductive methods. A number of the illustrations already cited in this section also involve counting. The answer to the numerous questions involving "How many...?" or "How much...?" are directly or indirectly dependent upon counting or measuring. For example, the answer to the question, "How many pupils are present in our class today?" may be found by counting.

Such statements as "There are thirty-six pupils here today," have little or no predictive value, however. As was indicated by Searles[8], this statement is merely a summary or descriptive statement. Nevertheless, such summary statements if collected over an extended period of time may provide the particular premises for an inductive proposition. For example, at one school the lunchroom supervisor had the job of ordering an adequate supply of milk to serve the pupils of a school with an enrollment of 391 pupils. She wanted as much milk as the youngsters would drink, but she did not want to order too much for it would spoil if kept too long. At first she ordered milk by guessing how much she thought the children would drink. She then kept a record of the number of bottles of milk consumed daily for two weeks. As a result of her study of the data collected, this lunchroom supervisor now has a standard order with the milk company for 408 half-pints of milk each school day.

This illustration can easily be extended to the problem which the milk company has in regulating its supply of milk to meet the needs of its multitude of customers. One of the important basic problems of business and industry, of course, is this problem of supply and demand. Millions of dollars are gained or lost in business and industry by predictions based on countable evidence gathered in past years. Obviously, only probable inferences are possible in these situations for unknown factors such as unexpected economic prosperity or depression, war, new inventions or products, and the like may alter the actual use or consumption of products.

Another familiar example involves the prediction of population trends. Before World War II, statisticians who dealt with population trends were predicting on the basis of census data of previous decades that the population of the United States was approaching a plateau, an expected limit. However, the so-called *war babies* that have been born since 1944 and other factors have sent the population soaring to over 170 million and the predicted plateau has been by passed.

Five ideas have been illustrated in this section as bases for probable inference. The suggestive subtitles were (1) Recognition of Authority, (2) Seeing Is Believing, (3) Measurement, (4) The Pragmatic Method, and (5) Counting. These methods are not distinct. In practical decision

making these methods may be intertwined and interconnected in such a way that it might be difficult to distinguish them one from the other as they are used by any given individual. Nor is their use limited to elementary and secondary school pupils. Adults, too, are dependent upon the senses of *seeing, hearing, feeling, tasting,* and *smelling* for information that may lead or help to lead to the solution of problems that confront them.

We shall now proceed to more refined and mature methods of probable inference.

TYPES OF PROBABLE INFERENCE USEFUL IN THE JUNIOR AND SENIOR HIGH SCHOOL

Even though generalizations reached through an examination of particular instances cannot be classified as certainties (the statistician would say that they cannot be assigned *probability one*), it has been pointed out that people in the workaday world depend heavily upon probable inference. In this section we shall continue our discussion of methods that lead to convincing men of the rightness of their generalizations. Five methods will be discussed: (1) The Method of Simple Enumeration, (2) The Method of Analogy, (3) Extending a Pattern of Thought, (4) Hunches and Their Relation to Proof, and (5) Testing the Hunches.

The Method of Simple Enumeration

Many of our beliefs have their roots in simple observations we have made over a period of time. For example, an elementary school youngster states that all snow is white. This is not only a summary of this pupil's past experience, but also represents a prediction that any snow which falls to the earth in the future will also be white.

This is a method used by budding young scientists in their general science classes as they heat ethyl alcohol in an open container to find the temperature at which it will boil. It was just through such a procedure that scientists arrived at such an accepted truth as "In an open container at sea level, sulfer boils at 445°C." Unfortunately, the students in our secondary school classes too often know the expected outcome of their experiments before they begin—they attempt to *confirm* something they already know. Furthermore, too frequently they perform one experiment and generalize from this insufficient data.

A television commercial might proclaim the results of a scientific test in the following manner:

Four hundred New York women in a recent test *proved* that Smith's lotion can prevent detergent burn. These four hundred New York women soaked their

hands for 30 minutes each day for two weeks in water containing a well known detergent. Their left hands on which they had used *no* Smith's lotion became rough, red, and sore. Their right hands, protected by rubbing on Smith's lotion, were soft and smooth and lovely. Buy Smith's lotion today and avoid detergent burn.

What is there in the fact that Smith's lotion has prevented detergent burn on the right hands of 400 New York women that would interest other ladies who saw and heard this advertisement? If we assume that the statements in the commercial are accurate, does this represent proof in the sense of a necessary inference or a probable inference? Such judgments must be made daily by every reader, radio listener, and television viewer as they are bombarded by propaganda not only from advertisements but also from political speeches and other arguments designed to influence public opinion.

Schaaf describes how the members of a class in ninth grade algebra studied their symbolic solutions of story problems involving signed numbers and then developed their own rules for the addition of positive and negative numbers. (They had used a number scale to find their initial answers.) He also reports that in formulating such a rule, there is ample opportunity for the students to realize the necessity of considering samples of all possible types of addition problems, how one contradictory case can destroy a universal conclusion, and the dangers of hasty generalization.[7] This method could easily be extended to determining a rule for the subtraction of signed numbers.

Generalizations based on simple enumeration are assertions made from a number of observed instances which are assumed to be representative but *not* assumed to be all of the instances of a given class. They are also assumed to have predictive power; that is to say, if you were to boil water in an open container tomorrow someplace along the seashore, you could reasonably expect the water to boil at 212°F. Certainly this experiment, if you perform it, could not possibly have been included as one of the instances on which the scientific generalization or proposition was based. Obviously, universal propositions, that is, propositions which begin "All ..." or "No...," can be destroyed by one contradictory case.

The hand lotion commercial is interesting because the viewer-listener is led to make the general conclusion that *all human hands are protected from detergent burn by Smith's lotion.* This then becomes the major premise for a syllogism. (Syllogisms will be discussed in the section on deduction or necessary inference.)

All human hands are protected from detergent burn by rubbing on Smith's lotion. (Major premise.)

My hands are human. (*Minor premise.*)

My hands are (can be) protected from detergent burn by rubbing on Smith's lotion. (Conclusion.)

This conclusion represents a necessary inference; however, nothing is ever implied in a conclusion that is not already implied in the premises on which it is based. The major premise in this case is a probable inference for it is based on the empirical evidence supplied by observing the 400 New York women. What effect then does this have on our conclusion? It simply means that we must recognize that the conclusion has only probability in its favor despite the fact that it was reached through necessary inference.

You will recognize the following generalizations that have been made and verified by simple enumeration: (1) water at sea level pressure freezes at 0°C. or 32°F., (2) all ravens are black, and (3) iron filings will cling to all magnets. Other generalizations that have been established as false because contradictory instances have been found include: (1) all swans are white, (2) all redheads have bad tempers, and (3) matter *can neither be created nor destroyed.*

Euclid must have been convinced of the validity of his conjecture that there is an infinite number of prime numbers by the never-ending string of them that he found by testing larger and larger numbers to determine whether or not they were prime. Euclid's proof in his *Elements* that *the number of prime numbers is infinite* is still admired for its simplicity and beauty.

As a mathematical recreation, one junior high school teacher asked his class if they could find a relationship between the even numbers greater than 4 and the sums of two odd prime numbers. Here are some of the relations that these pupils discovered:

$6 = 3 + 3$	$16 = 3 + 13 = 5 + 11$
$8 = 3 + 5$	$18 = 5 + 13 = 7 + 11$
$10 = 3 + 7 = 5 + 5$	$20 = 3 + 17 = 7 + 13$
$12 = 5 + 7$	. . . .
$14 = 3 + 11 = 7 + 7$	$30 = 7 + 23 = 11 + 19 = 13 + 17$

After working a while on this problem, one youngster exclaimed, "This could go on forever. There is no end to it." Is there an end? Is there an even number greater than 4 whose sum *cannot* be expressed as the sum of two odd primes? About 200 years ago a mathematician by the name of Goldbach proposed a proposition that may be stated in this way: *Any even number greater than 4 is the sum of two odd primes.* It has never

been proved. Prove this proposition and you will write your name indelibly into the history of mathematics.

Many of our beliefs have come into being as a result of using this method of simple enumeration. Its techniques are invaluable to the scientist whether he be a social scientist, biological scientist, or physical scientist. The method as exemplified by the illustrations in the previous paragraphs also serves the mathematician as a source of promising conjectures.

The Method of Analogy

"Analogy seems to have a share in all discoveries, but in some it has the lion's share."[4] The teacher of demonstrative geometry is familiar with the many similarities that exist between plane geometry and solid geometry. For example, the propositions related to parallel lines in plane geometry have their counterparts in solid geometry in the nest of propositions that concern two parallel planes intersected by a transversal plane. Analogies between the two subjects facilitate learning, provide a hierarchy of relationships that are more easily remembered, and suggest many conjectures that may later be proved or disproved.

In plane geometry there is an infinite number of regular polygons (i.e., polygons that have equal angles and equal sides). Using the method of analogy pupils readily generalize, therefore, that there must be an infinite number of regular convex polyhedrons (i.e., geometric solids with equal polyhedral angles and faces that are all congruent regular polygons). Further investigation of this conjecture reveals a surprising result. There are five and only five regular polyhedrons—the tetrahedron, the hexahedron, the octahedron, the dodecahedron, and the icosahedron. Analogies may be useful, but they do *not* substitute for proof. They provide only additional conjectures to be investigated.

In plane geometry pupils learn that the sum of the interior angles of a polygon increases as the number of sides increases. Suppose that the sides are extended successively to form exterior angles as shown in Figure 2. How does the sum of these exterior angles change as the

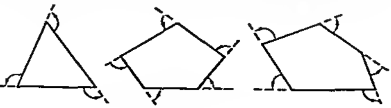


FIG. 2

number of sides of the polygon is increased? The youth who uses the method of analogy is going to be misled, for in any convex polygon, the sum of the exterior angles formed by extending the sides successively is a constant. The sum is 360° regardless of the number of sides.

In algebra, students learn to factor such expressions as $x^2 + 5x + 6$, $x^2 + 6x + 5$, and $x^2 + 6x + 9$. If the expressions are of the form $x^2 + bx + c$ they readily learn the principles that c is the product of the second terms of the binomial factors and that b is equal to the sum of the factors of c. By analogy the pupils can apply these ideas to factoring $a^2 + 5a + 6$, $m^2 + 6m + 5$, and $y^2 + 6y + 9$. It is relatively easy to factor this second group of examples if one knows how to factor the first set, because there are likenesses in essential details. By analogy, the same principles hold for examples such as $x^2 - 8x + 12$, $p^2 - 2p - 24$, $b^2 + b - 12$, and $w^2 - 16$.

Some of the most influential hypotheses of this scientific age have their origins in man's ability to use analogous reasoning. Niels Bohr proposed that the atom was similar to our solar system. Physics teachers may explain some of the intricacies of electrical theory by using the analogy of running water. The wave theory of light may be related to the rippling waves emanating from the spot where a stone is dropped into an otherwise placid pond of water. Polya describes[9] how reasoning by analogy led Euler to the solution of the problem of finding the sum of the infinite series of the reciprocals of the squares.

$$1 + \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{16} + \tfrac{1}{25} + \tfrac{1}{36} + \cdots$$

It is interesting to note that Euler's solution of the problem violated the accepted procedures held by him and his contemporaries for finding the sums of the finite series with which they had worked. Sometimes it takes bold, unorthodox methods to solve problems that resist attacks by standard methods.

Extending a Pattern of Thought

Great baseball players have become almost legendary figures in the history of the New York Yankee professional team. In the spring of 1951 a promising young rookie whose name was Mickey Mantle was taken to the spring training camp. It was hoped that Mantle would continue a succession of Yankee *greats* that included Babe Ruth, Lou Gehrig, and Joe DiMaggio. One rabid fan made the following observation:

> Babe Ruth, a great Yankee baseball player, wore the number *3* on his uniform.
> Lou Gehrig, a great Yankee baseball player, wore the number *4* on his uniform.

Joe DiMaggio, a great Yankee baseball player, wore the number *5* on his uniform.

Mickey Mantle should wear the number *6* on his uniform. (He does wear 7.)

The obvious prediction made by this hopeful fan from the data presented above was that *Mickey Mantle would become a great baseball player*. A more general statement is implied by this prediction, namely, *New York Yankee baseball players are great if the numbers on their uniforms are consecutive with past great New York Yankee baseball players*. This fan's prediction has come true, for Mickey Mantle has become a great baseball player. Does this then represent a verification or a disproof of the generalization concerning Yankee players? Who will wear Yankee uniform number *6*? Is he destined to be great by virtue of the number of his uniform?

It seems that man possesses an innate capacity to see relationships and order in the events which he observes. In the illustration above, a baseball fan observed a certain continuity of form. Few if any of us would grant that the numbers of the uniforms worn by baseball players had anything to do with their greatness, yet this example does illustrate a method of discovery that often results in fruitful conjectures. One might look upon this procedure as a special kind of analogy.

This is the method used by Edmund Halley, the English astronomer, when he predicted the appearance of the comet which now bears his name. Halley, the first great calculator of comet orbits, studied the paths of twenty-four comets. These represented all the comets for which he could find accurate observations. Twenty-one of the comets appeared from outer space and left the areas of their observers by different paths. Three of the comets appeared to follow the same orbit. Could these three comets really be sources of reports of the same comet revisiting the earth? he asked himself. If so, then its orbit must be an ellipse. It also seemed reasonable to assume that if this were the same celestial body, then it should return to the visibility of observers on the earth at approximately the same time intervals. Reports indicated that the three comets had been nearest the sun on August 24, 1531; October 16, 1607 (an interval of 76.2 years); September 24, 1682 (an interval of 74.9 years). Hence, Halley concluded that the comet would complete one revolution in its orbit about every seventy-five or seventy-six years. On the basis of this evidence he finally decided that the reports represented observations of the same comet, and he predicted the comet's return.

An analysis of Halley's methods in the development of his theory concerning the comet which now bears his name reveals that he used a variety of inductive procedures to arrive at his generalization. He noted,

for instance, that the comets of 1531, 1607, and 1682 had similar orbits as they flew across the heavens in their majestic ways. This is reasoning by analogy. On the basis of these three analogous cases that he had enumerated be made the prediction that the comet would return in 1758 or 1759. Several years after his death, the comet actually did return, in March 1759! The comet also reappeared in 1835 and 1910. Its next appearance is predicted for 1985 or 1986.

The dominating feature of Halley's discovery is not the analogy but the striking feature is his extrapolation, the act of going beyond the data to predict the return of the comet. It is his logical extension of the pattern in which he found the instances that makes this illustration different from those cited in connection with the methods of simple enumeration or analogy.

Three English astronomers have traced the reports of Halley's comet as far back as 240 B.C. These men found records of what are assumed to be twenty-nine appearances of the comet. The intervals between its successive appearances have varied from 74.5 to 79 years. This variation is due to the disturbing actions of the gravitational attractions of the planets of our solar system as the comet moves about in its elliptical orbit.[10]

In elementary arithmetic children generalize on the basis of their experiences that the subtrahend in subtraction can never exceed the minuend. From this generalization it follows logically that the difference in a subtraction example can never be greater than the minuend. The mathematician is continuously seeking to *relax* such rules to gain more generality. Why can't the subtrahend be greater than the minuend? Let us suppose that it can. What then is the answer to the question, $7 - 8 = ?$ No natural number will answer it; a new number, the negative number, must be invented. Mathematicians have chosen the symbol -1 to express the answer to this question. With this invention they have opened up a whole new domain of numbers that are similar to the positive numbers already familiar to the child, for the negative number *makes subtraction possible even when the subtrahend is larger than the minuend.* The use of inductive sequences to lead students to discover for themselves the rules for operating with signed numbers was discussed on pages 47 to 49 of Chapter 2. It was also explained there that whereas the definition $(-2)(-3) = +6$ was in a sense completely arbitrary, it was, on the other hand, a necessary definition if we wished the distributive law to hold for negative as well as positive numbers. This *principle of permanence of form* when so used is a type of analogy.

This same principle provides the reason for defining fractional negative

and zero exponents as we do. Explanations of this rationale help students to understand better both the definitions themselves and the role of definition in a mathematical system. It should be noted however that none of these developments by analogy or definitions embodying extensions of form constitutes proof.

It is difficult and probably impossible to separate completely probable inference and necessary inference in many situations. For example, let us consider the geometric figures in Figure 3. There are several observa-

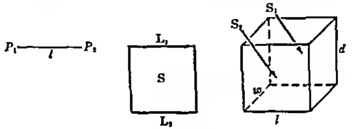


FIG. 3

tions that we can make about the relationships among the point, line segment, square, and cube. One of the most obvious is that here are geometric figures with zero, one, two, and three dimensions, respectively. The line segment, the one-dimensional figure (it has only length), starts with a point and ends with a point. We may then consider that the line segment is bounded by two points. Observe then that the one-dimensional figure is bounded by zero-dimensional figures.

The square, a two-dimensional figure, is bounded by four line segments. Hence, the two-dimensional figure is bounded by one-dimensional figures. Similarly, we observe that the cube which is a three-dimensional figure is bounded by six square faces. Hence, we arrive at the generalization that each of these geometric figures is bounded by the figure with one less dimension. What would this mean if we were to extend our thinking to the fourth dimension? Does it not seem reasonable to assume that the corresponding fourth-dimensional figure would be bounded by cubes? By going back, analyzing the process by which a point might be regarded as *generating* the line segment; the line segment, the square; and the square, the cube; one can find a simple process and derive an algebraic formula for predicting the number of elements, (points, lines, planes) in each newly formed configuration. From the three-dimensional cube, we can extend the process by analogy to find the points, lines, planes, volumes (cubes), and *content* of the four-dimensional tesseract and even on to higher dimensional configurations. We haven't the space

to do this here, but there are several sources which make the process clear enough to be used with high school algebra and geometry students.[11]

The point is that mathematicians, even before the physicists had need for the concept, pioneered the conception of a fourth dimension, and that this pioneering owed much to the use of the idea of permanence of form and the extension and modification of old concepts as a source of fruitful conjectures and assumptions on which to base their deductive proof in the extended geometry.

The illustrations which have been cited in this section up to now have dealt with what is commonly called *extrapolation*. Extrapolation implies the extension of ideas beyond the known or observed instances.

Wertheimer in describing a discovery by Galileo indicates how he used the method of extension of form to fill in a gap in his theory of motion. Galileo had been trying to determine whether or not free falling bodies obey some general law that could be easily understood by man. The instruments with which he had to work were rather crude, and he was having difficulty conducting his experiments. Finally he hit upon the idea of using an inclined plane and considering a vertical descent as merely a special case of the inclined plane, the case of falling at an incline of 90°. In the course of his experiments Galileo noted that as spheres rolled *down* an inclined plane, their velocities became greater and greater (positive acceleration). As spheres were rolled *up* an inclined plane, their velocities became less and less (negative acceleration). But what about motion on a horizontal plane? Galileo concluded that the motion of a sphere must have an acceleration of zero on a horizontal plane. A modern physicist might express the concept in this way: "If a body moves horizontally in a given direction, it will continue to move at a constant velocity for all eternity, if no external force changes its state of motion."[12] The key factor in making this generalization is the extension of form. This time it is a gap filling process; it is *interpolation*.

It should be noted that the process of extension of form is merely a logical extension of known and accepted relationships. It approaches the formality of deductive inference, but it does not carry the authority, the necessity of the implication, that characterizes deduction.

Hunches and Their Relation to Proof

Hypotheses, theories, and laws based on experiential data are the core of the sciences. They are generalizations fundamentally dependent upon probable inference from observable evidence. In mathematics the generalizations that constitute the core of the discipline are the results of rigorous proof. The test of these deduced theorems in mathematics

does not reside necessarily in its correspondence with man's interpretation of his experience although, as we shall see, experience may help him in rationalizing certain elementary principles of mathematics. The test of Euclidean geometry, for example, is in its logical consistency; that is, the noncontradictory nature of the theorems within the system. If contradictory theorems are validly deduced in a logical system such as Euclidean geometry, then the logical foundations of the system must be examined for inconsistent assumptions.

One might reason then that conjectures and inductive reasoning have no relation to the deductive science of mathematics. But Polya points out that although, "We secure our mathematical knowledge by demonstrative reasoning, ... we support our conjectures by plausible reasoning."[13] For example, consider the elementary school child who was struggling with the addition of $\frac{1}{3}$ and $\frac{1}{2}$. The youngster insisted that the answer was $\frac{2}{5}$ (an all too common answer). The teacher was a patient and understanding woman who suggested, "Let's check your answer using our fractional parts." (These were circles cut into various fractional parts.) The child apparently had little conception of what a fraction was, for he seemed surprised that two fifths was less than one half. After a short period of experimentation with the circles cut into fractional parts, the child decided that the answer should have been five sixths. The process of changing the two fractions to sixths was also made meaningful using the fractional parts. This boy had attempted to support his conjecture by plausible reasoning. When it failed to support his answer he had to find a new conjecture. Ultimately he got a solution using the well-known algorithm.

An intermediate algebra class was studying the quadratic function which is determined by $f(x) = ax^2 + bx + c$. They first studied the effects on the graph of the function as changes were made in each one of the parameters a, b, and c while the other two were held constant. The class was also asked to find graphically the real roots, x_1 and x_2, where they exist when the function is equal to zero. A table of data was then made from the results of different students' graphs (Table 1).

In a discussion following the organization of the first five columns, these four questions were posed:

1. What is the value of $x_1 \cdot x_2$? Record your answer in the table of data.

2. Is there any relation between the product of the roots of $ax^2 + bx + c = 0$ and the arbitrary constants a, b, and c?

3. What is the value of $x_1 + x_2$? Record your answer in the table.

4. Is there any relation between the sum of the roots of $ax^2 + bx + c = 0$ and the values of a, b, and c?

TABLE 1

No.	a	b	c	x_1	x_2	$x_1 \cdot x_2$	$x_1 + x_2$
1	1	−7	12	4	3	12	7
2	1	5	5	−3.4	−1.2	4.1	−4.6
3	1	4	4	−2	−2	4	−4
4	1	4	−5	−5	1	−5	−4
5	2	2	−5	2	.5	1	2.5
6	1	6	−3	.50	−6.5	−3.3	−6
7	3	6	8	.9	−3	−2.7	−2.1
8	1	7	6	−1	−6	6	−7
9	4	0	−1	.5	−.5	−.25	0

After a brief look at the table of data one boy ventured this guess, "The product of the roots is equal to c." This was called the "Garrett Conjecture," naming it after the boy who had made the guess.

"That seems to be true at least approximately, for 1, 2, 3, 4, 6, and 8, but look at 5, 7, and 9," one of the girls suggested.

The "Garrett Conjecture" was discarded after consideration of this young lady's statement, and the students were encouraged by the teacher *to look for likenesses and differences among the two distinct groups of equations stated above.* After two other false starts the "Livingston Conjecture" that "$x_1 \cdot x_2 = c/a$" was proposed. This gained considerable support, although not 100 per cent, apparently because of Examples 6 and 7.

A similar procedure was used to reach the conjecture that $x_1 + x_2 = -b/a$. The discussion followed a pattern that was analogous to the one related to the product of the roots.

"Can we prove these relations?" someone asked.

The problem of proving the relations was assigned as a part of the homework with the caution: "We may not know enough yet to answer this question. If we don't, then suggest what further information you think is needed." During class discussion the next day several members of the class indicated that they were sure that if the general equation $ax^2 + bx + c = 0$ could be solved for x, then the conjecture concerning the relations of the roots could be proved or disproved. Some of the students had attempted to solve the general equation by guessing, then checking by substitution and by factoring, but they had failed to find a solution. The students were then taught *how to solve quadratic equations by completing the square.* This method, after it was mastered, enabled the group to find the solution for the general quadratic function above and hence *to proceed to prove their conjectures about the product and sum of the roots.*

some negative numbers. For simplicity you could start at $n = -1, -2$, and so on. Here you will find additional support for the conjecture that $n^2 - n + 41$ will always produce a prime number. If you think you have enough evidence, you are doomed to disappointment for when $n = 41$, a little simple factoring will indicate to you that $41^2 - 41 + 41$ is divisible by 41 for $41^2 - 41 + 41 = 41\,(41 - 1 + 1) = 41 \cdot 41$. It should now be evident that no expression of the form $ax^2 + bx + c$ will *always* yield a prime number, for when $x = c$ $(c \neq 0)$, the number is divisible by c, hence, it is not prime.

On the basis of empirical evidence alone we can never be sure whether or not we have made a hasty generalization. The statement that *all swans are white* was verified time and again before explorers had reached Australia. However, on this continent in the Pacific, black swans were found. This discovery, of course, falsified the generalization. Similarly we saw in the above illustration that $n^2 - n + 41$ yields a prime number for 80 consecutive integral values of n, yet for the eighty-first value the expression is *not* a prime!

The illustrations cited above indicate an important point often overlooked by the elementary and secondary teachers of mathematics, namely, that we need to consider not only the verification but also the falsification of any given proposition. Consider this proposition: "If the difference between the sum of the odd-numbered digits and the sum of the even-numbered digits of an integer is divisible by 11, then the integer is divisible by 11." One member of an algebra class studying proofs for various tests of divisibility had selected the number 665,379 to test the generalization. Since $\dfrac{(9 + 3 + 6) - (7 + 5 + 6)}{11}$ equals 0 and by division he found that 665,379 was in fact divisible by 11, this youngster added his datum to the mounting evidence collected by the class verifying the original proposition. In an attempt to extend this generalization, he also proclaimed that by replacing 11 by 13 in the proposition he had a new conjecture which appeared to be true. This proposition was quickly falsified however with evidence presented by several members of the class. It is worthy to note two things about this illustration. First, regardless of the mass of evidence gathered to support the first proposition, it is not proved by accumulation of such data. It still remains to be proved deductively. In the case of the second proposition, only one contradictory case was needed to disprove it. This method of disproof is by counterexample which is discussed more thoroughly in the section of this chapter dealing with necessary inference.

There are propositions also where evidence may be collected to verify

them completely or to falsify them as the case may be. For example, let us consider the conjecture that *six is a perfect number.* (A perfect number is a natural number which is the sum of all its divisors, except itself, which are natural numbers.) Since the only natural numbers which are divisors of six (excluding itself) are 1, 2, and 3 and $1 + 2 + 3 = 6$, then six is a perfect number. Suppose it is proposed that *eight is a perfect number.* The divisors of eight to be considered are 1, 2, and 4. Since $1 + 2 + 4 \neq 8$, then eight is *not* a perfect number and the proposition is false. Note in particular that the propositions concerning perfect numbers could be labeled true or false on the basis of collected data because they dealt with a finite universe, whereas the proposition concerning the divisibility by 11 encompasses an infinite set, the set of all integers. Although the proposition concerning divisibility by 11 can never be completely verified by the collection of confirmatory instances, it can be proved by the mathematician and even by high school students. This again demonstrates the power of mathematical techniques.

Many times we are confronted with a conjecture of what a mathematician would characterize as an *existence theorem.* For example, someone might propose "There are black roses" or "There is a real number x such that $|x| + 3 = 2$." If a bush having black roses is found, then the first statement is verified. However, if one has not been found up to this time, this fact is no assurance one will not be found within the next hour or within the next day. As for the other proposition, the search for a solution by *trial and error* might continue for a long time because it can be proved that there is no real number which when substituted for x will make the sentence $|x| + 3 = 2$ a true statement. Hence, we say that the solution set for this equation is a null or empty set.

Extending the illustration concerning perfect numbers, some mathematicians have proposed that "There are *no* odd perfect numbers." One way for you to become famous, among mathematicians at least, is to find a perfect number that is an odd number. This is a rather interesting proposition, because the counterexample, if found, disproving the proposition would also establish an existence theorem; that is, that there exists an odd perfect number.

One adolescent, as he looked at several prime numbers including 11, 13, 17, 19, and 23 among others, exclaimed, "All prime numbers are odd numbers." Is this a true statement? Obviously it is false, for, as one of his classmates quickly pointed out, 2 is an *even* prime number. This and several of the other illustrations point out that propositions using the quantifiers *all* or *no* are falsifiable (by counterexample) but may

not be verifiable. Even though thousands of white swans had been seen adding credence to the proposition that *all swans are white*, one black swan seen in Australia was sufficient to destroy the generalization.

In elementary and junior high school mathematics we sometimes experience some difficulties using an inductive approach to develop a mathematical concept if measurement is involved. For example, members of a junior high school mathematics class were measuring the interior angles of various sizes and shapes of triangles to find the sum of the angles of each type. Their teacher had placed three boxes on her desk labeled: "1. Triangles—sum of angles less than 180°," "2. Triangles—sum of angles greater than 180°," and "3. Triangles—sum of angles equals 180°." The pupils were industriously measuring angles and sorting their triangles. After a short while each box had some triangles in it. At the end of this exercise could these boys and girls draw any valid conclusion from the data which they had collected? They could if they understood the approximate nature of measurement and if they used an average (the mean) of the sums of the interior angles. All of the sums, after the measurements of the angles had been rechecked, fell within the range from 178° to 183° inclusive. The mean, to the nearest degree, was 180°, and the class readily accepted the conclusion that *the sum of the interior angles of any triangle is 180°*. Variations from this mean were attributed to small errors in measurement resulting from poor estimates in reading protractors, from inaccuracies in the protractors themselves, from the careless construction of the triangles, and from the approximate nature of measurement.

The leap from particular cases, sometimes called inductive premises, to a universal conclusion is not a deductive process. However, the reverse step of applying the generalization to particular situations is deductive. For example, a primary school child may, through a series of experiences, note that three apples and two apples are five apples, that three cookies and two cookies are five cookies, that three children and two children are five children, and conclude that three of any class of objects and two more of the same things are five of the elements being considered. From this universal conclusion it follows that a group of three balls combined with a group of two balls will form a new group of five balls. The child can readily verify this deductive conclusion by counting. Note that the inductive premises on which the child based his generalization that $3 + 2 = 5$ now become deducible by necessary inference from the general statement.

Now it seems true that the inductive premises on which a generalized statement is based necessarily follow from the generalization. In a case

as simple as the one above, indeed it is obvious. Although it may seem strange to you, we may not always know all of the premises on which a given scientific theory or law is based. This is especially true of complex theories such as Einstein's Theory of Relativity. A theory may well be just an educated guess made from scanty evidence as compared to the wide range of phenomena subsumed by the theory or law. Not long ago an accepted theory in physics implied the existence of a particle called an antiproton. No one had ever seen such a particle but on the strength of the prediction from accepted theory, researchers set about the task of discovering the antiproton by experiment.

> One might suppose that the only way to discover a particle should be by experiment, but this is not so, although of course experiment is the judge of last resort. Sometimes theoretical physicists, from equations and calculations with pencil and paper, have predicted the existence of particles that have never been seen. These predictions, however strange some of them may seem, arise from a necessity to preserve basic principles which form the foundation of our present understanding of the physical universe. When necessary, physicists have been willing to entertain the existence of something never seen rather than imperil these firmly established foundations.[14]

The antiproton was not known or considered when the theory which implied it was devised. The verification of its existence, however, was essential to the verification of the theory.

Note that these authors consider experimentation, hence observational or empirical data, as *judge of last resort* in the physical sciences. (This is not true of necessary inference.) The experimental verification of the existence of the antiproton falls into place as one of the empirical premises for the general theory. Its discovery no doubt provides some measure of reassurance to physicists supporting the original theory. It should be noted, however, that although verification of deductive generalizations from a complex theory or law justifies the law's introduction, it provides no guarantee that it will hold in the future. It should be clear that if a deduced hypothesis is found to be false, then the logical bases of the theory or law must be re-examined.

Historically many of the rules of mathematics were made on the basis of empirical evidence. For example, the ancient Egyptians, as they surveyed the Nile Valley for the purpose of taxation after the annual flood of the Nile River, used the rule that the area of any triangle is equal to one half the product of its base and one of its sides. They did not specify what angle was to be included between these sides.[15] This same rule is often proposed by junior high school youngsters in their study of informal geometry. This method for finding the area of a triangle works fine

if the angle between the base and the side used is a right angle. It yields approximately correct results for the area of any triangle whose altitude and the side used, other than the base, are approximately equal. However, the critical examination of the area of an isosceles obtuse triangle whose base is the side opposite the obtuse angle will readily reveal the need for revising the rule of the Egyptians.

There are times when observations lead us to believe that a certain relation is true. However, we may be uncertain as to how this new relation fits into the general scheme of knowledge already accepted. The empirical rule of the Egyptians mentioned above is such a relation. The junior high school student is more convinced of the correctness of his guess that the area of a triangle is equal to one half the base times the altitude when he sees that two congruent triangles may be fitted together to form what appears to be a parallelogram with a base and altitude equal, respectively, to the base and altitude of the triangle. The area of one of the triangles then is seen to equal one half the area of the parallelogram. Thus, his conjecture appears to follow as a logical consequence of something he has already accepted.

Anytime that inductive generalizations may be interconnected in such a way that some of the conclusions may be deductively derived from the others, the whole structure becomes more convincing. There is little doubt that the psychological importance of such an interrelatedness of our beliefs is considerable; nevertheless, we need constantly to be reminded that this does not constitute proof in the sense of necessary inference.

It should be evident from the foregoing discussion that there is no such thing as absolute verification of an inductive generalization. Verification is only relative. Inductive generalizations are tentative judgments based on experiential evidence and are to be held *until further notice.* That is to say, if new and unfavorable evidence is accepted then the conjecture will be abandoned or modified to include the new evidence.

COMMON ERRORS IN THE USE OF PROBABLE INFERENCE

Before leaving the topic of probable inference or induction something should be said about the faulty use of the method. The entire range of human errors defies complete categorization, for the minds of men can err in what appears to be an infinite number of ways. We have referred to many of these errors incidentally already. The following statements represent a summary of some of the more common inductive fallacies.

Individual Prejudices and Preconceived Notions

All of us are influenced by personal factors in our attempts to collect data for the solution of a problem. This is especially true as it pertains to the collection and analysis of information on social issues and even on educational theory. We tend to accept evidence that supports our prejudices and forget the evidence against them. Although we often do not realize it, we see only what we want to see. Mathematics has a definite role to play in the impersonal collection and analysis of statistical data.

Strangely enough, individual prejudices and preconceived notions sometimes interfere in the learning of mathematics. Perhaps you have had pupils who were extremely reluctant to generalize rules or to accept rules for operations with negative numbers, the definitions of zero, fractional or negative exponents, or the trigonometric functions of general numbers. "Well it just doesn't seem right, and I don't think it should be that way. How can you take a number (the base) as a factor a negative number of times?" one boy remarked as he struggled with the meaning of 5^{-4}. The answer to his question is that you cannot take 5 as a factor negative four times. For negative numbers the phrase *exponent of the power* must be redefined. Do we teach some concepts in such a way that they must be unlearned when a pupil extends the idea? If so, can this be avoided?

Hasty Generalization—Jumping to Conclusions

What amount of positive evidence is necessary to reach a universal conclusion? This is a difficult question to answer. Eighty consecutive integers from -39 to $+40$ inclusive indicate that when n in the expression is an integer, $n^2 - n + 41$ will *always* be a prime number. This is considerable evidence. Yet when 41 is substituted for n, the resulting number is not prime!

The mathematician would say that no amount of observed evidence will establish a universal proposition (with probability 1). There is strong evidence that Goldbach was correct when he presented his conjecture that any even number greater than four can be expressed as the sum of two odd prime numbers. This conjecture will not be included among the accepted theorems of number theory, however, until it is proved deductively.

Nevertheless, we must not lose sight of the fact that in practical matters of everyday living, decisions must often be made on the basis of inconclusive evidence, at least on evidence that would not provide positive assurance (probability 1). These data may be a random sample

of the total evidence relevant to a problem as in the case of the statistical study of the polio vaccine. On the other hand it may be the marshalling in a court of law of all the evidence available, although known to be incomplete, in an attempt to convict a suspect of some crime. Someone has said that seldom do we have proof in the deductive sense in a case of law; most people are convicted on the basis of a high probability that they are guilty.

In the past we have talked so much about the dangers of hasty generalization that we may have overlooked the fact that one of the signs of genius is the ability to generalize correctly on one or a few cases. Perhaps we should be teaching children to generalize as soon as possible but to check the predictive value of their generalizations thoroughly before they accept them as anything more than tentative. When possible they should prove their conjectures and remove all doubt.

Begging the Question

The story is told of a check forger who walked up to a new and naive cashier to cash a personal check. When asked for identification, the forger pulled out his wallet and showed the cashier a picture of himself. "Yes sir, that's you all right," said the teller. "Here is your thirty-five dollars. Next." This is a clever and subtle way of arguing in a circle.

Those of you who teach demonstrative geometry will recognize circular reasoning as one of the pitfalls in proving originals. Perhaps for one or two arguments or statements the pupil will inadvertently assume what he is really trying to prove. The rest of his arguments may be all right; however, the proof is fallacious because the pupil has reasoned in a circle in a part of his proof.

False Analogy

In the section on the *Method of Analogy* we pointed out how pupils in geometry may reason fallaciously that there is an infinite number of regular polyhedra because there is an infinite number of regular polygons. It is also a false analogy to assume that, because Mary and Bill both have pleasing personalities and above average intelligence, Bill will be a conscientious mathematics student because Mary is such a student.

PROBABLE INFERENCE IN MATHEMATICS AND THE TEACHING OF MATHEMATICS

The point has been made previously that induction is *not* proof in the mature mathematical sense of the term *proof*. Nevertheless, probable inference *does* play a role in mathematics and in the teaching of mathe-

matics. Its role in mathematics is in the creation of new ideas such as (1) the statements of reasonable and fruitful postulates, (2) the discovery of conjectures requiring proof, (3) the discovery of proofs, and (4) the use of analogous reasoning in the study and development of new mathematics.

The role of probable inference in teaching is (1) to help students understand and appreciate the process of creation in mathematics (i.e., a student should have experience not only in proving conjectures of others but also in proving his own conjectures), and (2) to help teach for meaning and understanding. Inductive, informal, intuitive reasoning characterizes an early stage of the process of growth and maturation which culminates in a mature understanding of both probable inference and necessary inference, their interdependence, and their likenesses and differences.

NECESSARY INFERENCE

We now turn our attention to necessary inference—described by some authors as *the mathematical method* or the mathematical mode of thought.

Two questions usually arise when a person passes from one or more sentences, called *reasons* or evidence, to another sentence, called a *conclusion*. One of these is whether or not the sentences are true. The nature of the reasoning involved in answering this question for certain kinds of sentences has been discussed in the first part of this chapter. A second question is whether or not the conclusion follows necessarily from the sentences advanced as reasons (evidence) for it.

There are occasions when a person presents *some* reasons for a sentence being true but does not claim that the reasons are conclusive. He will admit that the conclusion does not necessarily follow from these reasons. The first part of this chapter has discussed this kind of reasoning. There are other occasions when he feels that the reasons he offers are conclusive and that the conclusion necessarily follows from them. We shall now turn our attention to this kind of reasoning and to the knowledge which can be taught to help pupils decide whether or not in an argument which purports to be conclusive the conclusion necessarily follows from the reasons offered.

THE CONCEPT OF INFERENCE

A pupil says, "We will need two more bottles of milk because two pupils don't have any." In this bit of reasoning the pupil is seeking to establish the truth of the sentence, 'We will need two more bottles of milk', by offering a reason, 'two pupils don't have any'. By accepting

the reason, one tends to accept the conclusion which is associated with the reason by the logical connector 'because'.

This instance of reasoning is an inference. An *inference*, sometimes called an argument, is a set of sentences some of which are regarded as providing evidence for the truth of another. We ordinarily speak of the latter as the *conclusion* and the sentences offered in support of the conclusion as *reasons* or *evidence*. A technical term used to denote a reason or evidence is *premise*.

Analyzing an inference—particularly an involved inference—requires that one be able to distinguish between reasons and conclusions. In doing this we are aided by certain logical connectors which often are used to introduce reasons. Among these are the following:

because	since
for	for the reason that
given	assuming
in as much as	in view of the fact that
may be deduced from	for example
may be inferred from	as shown by
on the hypothesis that	as indicated by.

Similarly, each of the following often serves to introduce a conclusion:

therefore	indicates that
hence	proves that
so	implies that
then	I conclude that
consequently	you can see that
we can deduce that	allows us to infer that
leads me to believe that	suggests that
it is obvious that	it follows that.

Consider the following involved inference and notice the use of linguistic cues which identify reasons and conclusions:

> Suppose we want to add $\frac{3}{4}$ and $\frac{1}{2}$ and write a simple name for the sum. We know the sum is not $\frac{3+1}{4+2} = \frac{4}{6}$ *because* $\frac{4}{6}$ is smaller than $\frac{3}{4}$. The sum of $\frac{3}{4}$ and $\frac{1}{2}$ must be larger than $\frac{3}{4}$. *Hence* we can't add numerators and denominators just as they are. What we do is find a common denominator. In this case 4 is a common denominator *since* it is exactly divisible both by 4 and by 2. We next divide the common denominator, 4, by 2 obtaining 2. We then multiply the number named by the numerator and the one named by the denominator of the fraction '$\frac{1}{2}$' by 2. The result is '$\frac{2}{4}$' which, we know, is another name for $\frac{1}{2}$. *Therefore* $\frac{2}{4} = \frac{1}{2}$. We can *therefore* substitute '$\frac{2}{4}$' for '$\frac{1}{2}$' in the name '$\frac{3}{4} + \frac{1}{2}$' and get '$\frac{3}{4} + \frac{2}{4}$'. A simpler name than '$\frac{3}{4} + \frac{2}{4}$' is '$\frac{5}{4}$' *so* our answer is $\frac{5}{4}$.

While this inference illustrates the use of signals for reasons and con-

clusions, it also shows that not every reason or conclusion is so identified. For example, the third sentence in the first paragraph is a reason, yet it is not introduced by any identifying expression. When identifying expressions are not used, one must decide from the context which sentences are reasons and which are conclusions.

The inference quoted above also serves to illustrate a point which is well known. Apart from logical connectors, e.g., 'because', 'hence', 'since', and 'therefore', used to signal a reason or a conclusion, there is no property in the form of a sentence which serves to identify it as a reason or a conclusion. In fact, the same sentence can function both as a reason and as a conclusion. An example of this double function is the sentence, "$\frac{2}{4} = \frac{1}{2}$." This sentence serves as a conclusion supported by the reason "... '$\frac{2}{4}$' ... is another name for $\frac{1}{2}$." Once it has been established, it then serves as a reason supporting the conclusion, "We can ... substitute '$\frac{2}{4}$' for '$\frac{1}{2}$' in '$\frac{3}{4} + \frac{1}{2}$'." The terms 'conclusion' and 'reason' are comparable to the terms 'teacher' and 'pupil' in that a given person may in one situation (context) be a teacher and in another situation be a pupil.

TEACHING ABOUT INFERENCES

Piaget and his collaborators[17, 18] have studied the language of children in an attempt to identify and analyze their reasoning. This research is helpful to a primary teacher in deciding what she can do in laying a basis for the concept of proof.

Piaget identifies an argument with verbal persuasion. The use of argument by children begins when they abandon physical force, threats, teasing, or name-calling to attain their ends. At first, argumentation is only a sequence of sentences, e.g., "I want a drink. Give me a drink." These may be related in the child's mind by something other than temporal order, but Piaget is not sure. The test he employed to identify the presence of logical relation was the use by the child of certain subordinate conjunctions. The first to appear was '*parce que*', our 'because' and subsequently '*puisque*', our 'since'. These were used to introduce a reason. He also found that these words were often used incorrectly. For example, a child might say, "The faucet is turned off because the water won't run" or "I had a bath because I am clean." He believed this inversion of cause and effect and hence reason and conclusion was caused by the juxtaposition of the two ideas in the child's mind with the 'because' functioning like 'and'.

Piaget found what is so apparent to parents and primary teachers—children's persistent asking *why*. What he realized, that some parents

and teachers do not realize, was that some 'whys' reflect only disappointment or frustration. The child really does not expect a reason; perhaps only consolation. If he is proffered a reason, he rejects it even if it is a good reason, perhaps by repeating, "But why?" or "But you haven't told me why." In the context of the present chapter, we are interested in the 'whys' that request explanation—the giving of reasons.

The child gradually learns even before he enters school that, according to the way we live, people justify their actions and beliefs and expect others to justify theirs. Sometimes this justification (reason giving) is supplied along with the statement or request being justified, e.g., "May I leave early because mother wants to take me downtown?" Sometimes it is supplied in responses to the question "Why?" or "How do you know?" Since such linguistic behavior occurs so frequently and naturally and is basically reasoning, here is where the primary teacher can tie in. She can tie in not only in teaching arithmetic but in all teaching, informal as well as formal. Through repeated uses of this kind of language, pupils can gradually be led to sharpen their concepts of a reason and a conclusion and the correct use of 'because', 'since', and 'if' to indicate reasons and 'then' and 'so' to indicate conclusions. Learning will be by imitation rather than by formulating and applying a verbal principle. The word 'conclusion' and its derivatives need not be taught. Teachers know expressions, e.g., 'What you think', that are better understood and serve well enough to name this idea. If 'reason' in "Give me a reason" is not understood by first graders, 'Tell me (us) why' or simply 'why' will serve the same function. The teacher can correct errors, e.g., "You should say, 'I am clean because I had a bath' not 'I had a bath because I am clean'," but probably without announcing a principle.

It is assumed that pupils will come to school with an understanding of the use of such words as 'all', 'some', 'not', 'each', 'every', 'true', and 'not true'. If their use of these words does not indicate an understanding, the teacher will have to teach the correct use. This teaching will probably proceed like that described above. But developing informally the concepts of reason and conclusion and the correct use of certain subordinating conjunctions represent the chief contribution of the primary teacher to the concept of proof. Reinforcement of the meaning of such words as those mentioned above can continue in the intermediate grades. This reinforcement will come as the terms are used in varied contexts. Additional words such as 'reason', 'conclusion', and 'therefore' can be added to the pupils' vocabularies as they increase in intellectual maturity.

Particularly helpful in teaching the pupil to reason is asking "Why?"

or "How do you know?" Piaget[19] believes that children arrive at answers to arithmetic problems which are correct or incorrect yet often do not know why they did what they did in obtaining the answer. An example he gives is the following: "When we ask: 'Why do you say 5?' of someone who has given the answer to the question: 'It takes 20 minutes to walk from here to x. Bicycling is 4 times as fast. How much is that?' the answer may be: 'Because I divided 20 by 4'. . . ." [20] This is description by the child; not explanation. The error is explainable, according to Piaget, in that, "The logical justification of a judgment takes place on a different plane from the invention of the judgment." [20] This seems to indicate that a teacher has to teach the child what a satisfactory reason is for this context. This can be done perhaps by asking, "But why did you divide 20 by 4?" or by saying, "You told me *what* you did. But *why* did you do that? Why didn't you *multiply* 20 by 4?" Gradually, the child's attention will be directed away from his action—what he did in obtaining the answer—and to the language of justification or explanation. He will learn unverbalized criteria of a *good* reason. Sharpening and verbalizing these criteria may be left for succeeding years.

As children become older and more mature, they will come to realize that reasons can be arranged in a chain in presenting a more involved argument. This realization may come as they persist in asking "Why?" An illustration of this might be the following exchange with the child speaking first.

"What are you doing?"

"I'm dusting the roses."

"Why?"

"Because I want to kill the aphids on them."

"Why do you want to do that?"

"Because the aphids spoil the flowers and we won't have any nice flowers to see and smell."

As the child experiences many instances of reasoning like this one, he probably comes to see that one can give reasons for reasons. And this essentially is the principle that a sentence may serve both as a reason and a conclusion.

THE CONCEPT OF NECESSARY INFERENCE

Let us consider some inferences. Suppose a pupil reasons as follows: "Most multiplication problems in our textbook result in a product which is larger than either of the two numbers which are multiplied. Since this problem in our book that I am doing is a multiplication problem, my

answer (product) should be larger than either of the two numbers I multiplied." Does his conclusion follow necessarily from the reasons he advances? Obviously not, for he might have been doing one of the few problems in the book in which one of the two numbers multiplied was a fraction. It is possible for his conclusion to be false even though both of his reasons (premises) are true.

Suppose the pupil reasons, "If a number is exactly divisible only by itself and one, it is a prime number. 43 is exactly divisible only by itself and one. Therefore, 43 is a prime number." Does his conclusion follow necessarily from the reasons he gives? The answer is 'yes', assuming the rules of reasoning we ordinarily use. And if the pupil had said, "43 is a prime number because it is exactly divisible only by itself and one," many people would say his conclusion, namely 43, is a prime number, necessarily follows from the reason he gave. At least, they would say that this is sound reasoning.

It was easy to decide in these instances whether or not the conclusions necessarily followed. But what of this bit of reasoning: '6 = 7' is a statement of equality because it contains an equals sign to the left and right of which are names of numbers. Does the conclusion, '6 = 7' is a statement of equality, necessarily follow from the reason supplied? This inference has the same form as the abbreviated one in the previous paragraph. Yet some of the same people who said the conclusion of the previous inference follows necessarily would say that the conclusion of the latter does not necessarily follow.

It appears we need a definition of 'necessarily follows' to settle this matter. However, instead of defining this term, let us define the term, 'necessary inference'. Then we can stipulate that the conclusion of a necessary inference necessarily follows from the premises. 'Necessary inference' can be defined in various ways. It is sometimes said to be an inference in which the conclusion follows inescapably from the reasons offered. It can also be defined as an inference which can be justified by laws of deductive logic. A definition which is both precise and useful in teaching is that a necessary inference is an inference in which it is not possible for the reasons to be true and the conclusion false. This is the one we shall use. Later we shall refer to the other two statements about a necessary inference and point out their use by a teacher in helping students discriminate between inferences which are necessary and those which are not.

A synonym for *necessary inference* is *valid inference*. An inference which is not necessary—whose conclusion does not necessarily follow—is called an *invalid inference*.

VALIDITY AND TRUTH

It needs to be kept in mind that validity is a property of an inference, but is not a property of a statement. Truth is a property of a statement, but is not a property of an inference. That is, we can correctly speak of an inference being valid or invalid and of a statement being true or false (but not both true and false). But we cannot correctly speak of an inference being true or false or a statement being valid or invalid. At this point someone may take an exception to what has just been said. He may point out that in responding to the inference, "Either an integer is an odd number or it is an even number. Since the integer 10 is not an odd number, I know that it is an even number," he is inclined to say, *true.* But the 'true' can be interpreted to be either a judgment concerning the truth of the conclusion, 'It (10) is an even number', or synonymous with, 'I agree that you have made a valid inference'. With such an interpretation—which seems reasonable if one ponders the matter a bit—the original statements about the relations among truth, statements, inferences, and validity still hold true.

There is one qualification which has to be made, however. The word 'inference' is used ambiguously. It is sometimes used to designate the set of reasons and conclusion. And it is sometimes used to designate only the conclusion. In the latter sense, one can speak of a *true inference* and thereby refer to the conclusion. Similarly, one can speak of a *valid conclusion* as a convenient way of referring to the validity of the inference of which the conclusion is a part.

To ascertain the relations between the validity of an inference and the truth of the sentences composing the inference, consider the following inferences. As you read the statements in each inference, make a decision about their truth. Then decide whether or not the inference is valid.

1. If the symbol for a whole number ends in '0', the number is exactly divisible by 10. '1000' ends in '0', therefore, 1000 is divisible by 10.

2. If the symbol for a whole number ends in '0', the number is exactly divisible by 3. If it is exactly divisible by 3, it is exactly divisible by 5. Hence, if the symbol for a whole number ends in '0', the number is exactly divisible by 5.

3. If a number is exactly divisible by 3, it is exactly divisible by 7. If it is exactly divisible by 7, it is exactly divisible by 5. Hence, if a number is exactly divisible by 3, the number is exactly divisible by 5.

4. If two angles of a triangle are congruent, two sides of the triangle are congruent. Two sides of an isosceles triangle are congruent. Therefore, two angles of an isosceles triangle are congruent.

5. If a geometric figure is a square, it is a circle. If a geometric figure is a triangle, it is a circle. Therefore, if a geometric figure is a square, it is a triangle.

6. The diagonal of square A divides the square into two congruent triangles. This is also true of the 100 other squares of various sizes I tested. Therefore, the diagonal of every square divides the square into two congruent triangles.

Let us now analyze these six inferences according to the truth of each reason and conclusion and according to the validity of the inference (Table 2).

TABLE 2

Inference	First Reason	Second Reason	Inference	Conclusion
1	True	True	Valid	True
2	False	False	Valid	True
3	False	False	Valid	False
4	True	True	Invalid	True
5	False	False	Invalid	False
6	True	True	Invalid	True

As can be seen from the table, an inference and its conclusion can be respectively: valid and true (1 and 2), valid and false (3), invalid and true (4 and 6), and invalid and false (5).

Inference 6 is the familiar probable inference. The inference could be rated as *good reasoning* or *sound reasoning* since these grading labels are more general and are used to rate probable and necessary inferences alike. But since it is logically possible for the reasons to be true and the conclusion false, the inference is invalid.

These examples show that a knowledge only of the truth of a conclusion is not helpful in determining the validity of the inference. A knowledge of the validity of an inference is by itself not helpful in determining the truth of the conclusion. Nor is a knowledge of the truth of reasons offered for a conclusion necessarily an indication that the conclusion is true. It is only when we know that the reasons are true and the inference is valid that we can know with certainty that the conclusion is true.

JUDGING INTUITIVELY THE VALIDITY OF AN INFERENCE

To be sure, pupils—and adults too—make judgments about the validity of inferences even though they know no patterns of valid inference. These judgments are made by intuition. This intuition tends to be based on the definition of a valid inference. If it is possible for

the conclusion to be false when the reasons are true, the inference is invalid. One kind of intuitive testing of the validity of an inference, then, consists of trying to find an instance in which the reasons are true and yet the conclusion is false. Suppose a pupil has always been successful in checking his addition problems by casting out nines. He might be inclined to say that if the problem checks by casting out nines, the sum is correct. Another pupil might point out the following problem which *checks* but whose sum is incorrect.

179	8
24	6
803	2
65	2
1062	0

The reason, i.e., the problem checks, is true but the conclusion, i.e., the sum is correct, is false. Hence, the inference is invalid.

This kind of testing is sometimes spoken of as finding a *counterexample.* The specific problem above is an example which counters or contradicts the conclusion stated. If one can find a counterexample, the inference is invalid. But if one cannot find a counterexample, the inference is not necessarily valid. It may be invalid, but the person does not know enough about the subject to find a counterexample. In intuitive testing, the tendency is to accept the inference as valid if no counterexample can be found. But the possibility of a counterexample provides the desire to be certain that this is not possible. This can be regarded as the motivation for a study of proof.

Related to the method of using a counterexample is the well-known principle of logic that to disprove a proposition of the form all A is B, one can prove its contradictory, some A is not B. To prove this, one needs to find at least one A which is not B. A child in responding to the statement, "All the dogs on our street are nice," may say, "No, Bruno is not nice. He barks at me and jumps on me." At a more advanced level, an author[22] of a trigonometry textbook writes:

> We shall now consider functions of an angle which is the sum of two angles. It seems reasonable to say that if the functions of 30° and 45° are known, the functions of 75° can be obtained. For instance, is sin 30° + sin 45° = sin 75°? Obviously this is false because $\frac{1}{2} + \frac{\sqrt{2}}{2} = .5 + .7 = 1.2$ which would make sin 75° greater than 1. This proves that, in general, $\sin (A + B) \neq \sin A + \sin B$. Likewise, sin 2$A$ is not identically equal to 2 sin A because $\sin 60° \neq 2 \sin 30°$.

In the same vein, to disprove a statement of the form, no A is B, one can prove the contradictory: some (at least one) A is B. To a pupil

who claims that no similar triangles are congruent, the teacher can exhibit two triangles XYZ and ABC, which have two angles and the included side of one congruent to two angles and the included side of the other.

Teachers can make use of the principle of counterexample when encouraging a pupil to test a generalization he has reached by induction. If he says or implies that dividing one number by another results in a quotient that is smaller than the dividend, the teacher can pose the problem $6 \div \frac{1}{2}$. Pupils will learn by imitation this method of counterexample to prove the falsity of certain statements. By the time they reach the seventh or eighth grade, they probably are mature enough to understand the principle involved if it is stated as: to disprove a statement that claims something is true of all members of a group, we need only to find one exception, in other words, a counterexample.

Another way of demonstrating intuitively the invalidity of an inference is by means of *logical analogy*. Suppose a pupil reasons, "If a polygon is equilateral and equiangular, it is a regular polygon. Polygon $ABCDE$ is a regular polygon. Thus, polygon $ABCDE$ is equilateral and equiangular." The pupil knows that all these statements are true so he claims that this is a valid inference. The teacher might reply, "According to your reasoning, the following inference is also valid: If a person is singing, he is alive. Tom is alive. Thus, Tom is singing. This conclusion, as you can see, does not necessarily follow. Yet this inference has the same form as the one you made. How can one be valid and the other invalid?"

This refutation by logical analogy is effective if the analogy is apparent to the person making the invalid inference and if the refuting inference is in terms of sentences easily decidable as true or false. Of course, if an inference is not refuted by logical analogy, this does not mean that the inference is valid. But as in using counterexamples, an inference intuitively is considered valid if a refutation by logical analogy cannot readily be found.

It is interesting to note that refutation by logical analogy can be subsumed under the technique of using a counterexample. According to this point of view, what is being tested is not the particular inference but the inference model or formula. If a particular inference fitting this model can be found whose reasons are true but whose conclusion is false, the inference model is an invalid model. Finding such a particular inference constitutes finding a counterexample. In the case of the inference above, the inference model may be regarded as: '*If p then q. q therefore p.*' We were able to find an example of this model, that is,

a particular inference, which was invalid. This served as a counterexample of the model. Hence, we know that by using this model and true reasons we cannot be sure that the conclusion is true. But more about inference models later.

MORE RIGOROUS TESTING OF THE VALIDITY OF INFERENCES

Consider the following reasoning:

> In arguing that a statement is true, the statement becomes the final conclusion in the argument. Now if this conclusion goes beyond any of the reasons used in the argument, it is invalid. But if it does not go beyond the reasons, it does not reveal any new knowledge, that is, knowledge not implicit in the reasons. The fact is that the conclusion either has to go beyond the reasons or not. Therefore, statements are either invalid or they do not reveal any new knowledge.

Does the conclusion of this involved inference necessarily follow from the reasons offered? Some people will say that it does and that the inference is valid. Others will say that it does not and that the inference is invalid. How shall we decide which group is correct? Intuitive judgment is of no help, for each group is convinced they are right. It is obvious we must have some rules to go by. One of the values of teaching students what it means to prove a statement is to enable them to move from a concept of a proof as *that which convinces me* to the concept of a proof as *that which satisfies laws of logic.* To put it another way, the aim is to teach the student when he can say 'therefore' and when he cannot. Without objective and generally accepted criteria, there is no way to settle a difference of opinion as to whether or not a statement has been proved. Instruction in the elementary and junior high grades paves the way for clinching the more sophisticated concept of a proof at the senior high and junior college levels.

At the senior high level, probably in plane geometry, the commonly accepted names, 'valid inference' and 'necessary inference', may be introduced at the same time the precise definition of such an inference is given, namely, an inference is valid if and only if it is not possible for the conclusion to be false when the reasons are true. This more precise concept of a valid inference seems to be useful before proceeding to teach a technique for judging the validity of inferences. We shall now turn to two general techniques for accomplishing this.

EULER CIRCLES

This technique, credited to the great Swiss mathematician, Leonhard Euler, can be used to test inferences composed of, or reducible to,

sentences of the form all A is B, no A is B, some A is B, and some A is not B. It consists of using circles (actually any closed geometric figure will do) drawn in various relative positions to visualize the relations asserted by these sentences.

Let us diagram the statement, 'All hexagons are polygons'. We shall use two circles, the interior of one to represent the set of hexagons and the interior of the other, the set of polygons. By drawing, as shown in Figure 4, the circle, whose interior represents the set of hexagons wholly

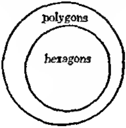


FIG. 4

within the one whose interior represents the set of polygons, we show that all hexagons are included in the set of polygons. The circles also depict the same proposition stated as 'A hexagon is a polygon' or in *if-then* language as 'for every geometric figure, if the figure is a hexagon, it is a polygon'.

All we know from the statement, 'all hexagons are polygons', is that the set of hexagons is included in the set of polygons. We do not know whether the part of the interior of the circle representing polygons but not hexagons has a member or not. Hence we cannot validly infer *from this representation* that some polygons are not hexagons, or not all polygons are hexagons. These statements, we know, are true. But they are so for reasons other than 'all hexagons are polygons'.

If we know that $ABCDEF$ is a hexagon, we can represent this statement as in Figure 5, and if we combine this statement with the previous one, we have (Fig. 6).

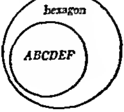
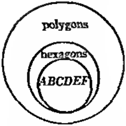


FIG. 5



All hexagons are polygons
ABCDEF is a hexagon
FIG. 6

Can we validly conclude that *ABCDEF* is a polygon? The diagram shows that we can. It is not possible for the conclusion to be false and the two reasons true.

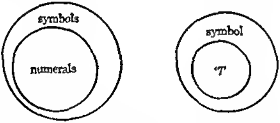
Is the following inference valid?

All numerals are symbols.

'7' is a symbol.

'7' is a numeral.

The conclusion certainly is true, but is the inference valid? To see, we diagram it (Fig. 7).



All numerals are symbols '7' is a symbol
FIG. 7

When we put these two diagrams together, we can do it in at least two ways (Fig. 8). Both diagrams represent correctly the two reasons.

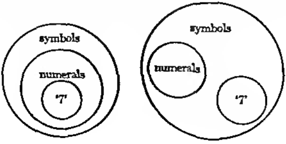


FIG. 8

Since there are two correct representations, there are two possible conclusions. The one given in the inference, therefore, is not unavoidable. We can escape the conclusion via the right-hand diagram which shows that whereas '7' is a symbol it is not necessarily a numeral.

Let us now consider the statement, 'No triangle has two right angles'. To see clearly how this relation can be represented by Euler Circles, we change the statement to 'No triangle is a figure having two right angles'. This appears to be equivalent to the given statement. This latter statement implies that the set of triangles is excluded from the set of figures having two right angles. We can visualize this relation by the diagram in Figure 9.

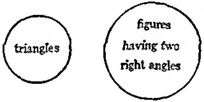


FIG. 9

From this diagram we can immediately draw an inference concerning the truth of these statements: (1) 'No figure having two right angles is a triangle, and (2) 'All triangles have two right angles'. The first is true; the second is false.

Is the following inference valid?

No even number is exactly divisible by 3.

14 is an even number.

∴ *14 is not exactly divisible by 3.*

To decide, let us represent each reason by Euler Circles, and then combine the circles. The two closed curves on the left in Figure 10 rep-

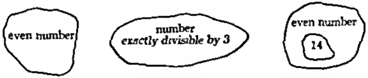


FIG. 10

resent 'No even number is exactly divisible by 3', and the right-hand diagram represents '14 is an even number'.

Combining the closed curves to represent both reasons, we have the diagram as shown in Figure 11. This diagram showing that the conclusion '14 is not divisible by 3' follows necessarily. The inference is valid even though the first reason is false.

FIG. 11

Is the following inference valid?
No triangle is a hexagon.
No hexagon is an octagon.
∴ No triangle is an octagon.

The inference sounds plausible perhaps because all the statements are true and they appear to be related by transitivity. But let us see (Fig. 12).

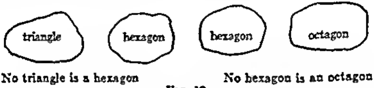


FIG. 12

When we combine the Euler curves so as to preserve these relations, we find we have three possibilities among others (Fig. 13).

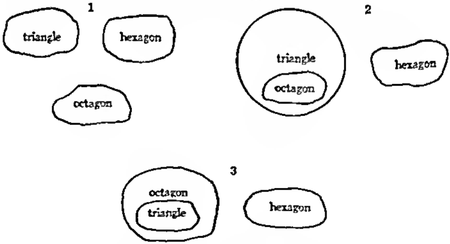


FIG. 13

There are, therefore, at least three possible conclusions: (1) no triangle is an octagon, (2) all octagons are triangles, and (3) all triangles are octagons. Hence, the one proposed in the inference does not follow necessarily.

Particular statements like 'Some algebraic equations have integral roots' and 'Some arithmetic problems are not hard' can be represented by Euler curves. In making the representation, we must remember that 'some' is always used in the sense of 'at least one'. 'Some equations' implies 'there is at least one equation' (Fig. 14).

Some equations have integral roots Some arithmetic problems are not hard
FIG. 14

The '*' indicates that there is at least one member in the class identified, i.e., at least one algebraic equation having integral roots and at least one arithmetic problem which is not hard.

Suppose someone argues, "Since some equations have integral roots, some do not have integral roots." Is this valid reasoning? It is not. The fact that the set consisting of things that are both equations and have integral roots is not empty, that is, has at least one member, gives us no information about whether the set consisting of things that are equations and do not have integral roots is empty or not. As a matter of fact, we know that there are equations which do not have integral roots but not for the reason given in this particular inference.

Which of the following statements follow necessarily from 'Some arithmetic problems are not hard'?

1. Some hard problems are not arithmetic problems.
2. Some hard problems are arithmetic problems.
3. Some arithmetic problems are hard.

The answer is that none does, for the reason, by itself, provides no information about whether the other sets represented in the overlapping Euler Circles are empty or not. These inferences illustrate again the point made earlier that a conclusion can be true but not because of the reason(s) given.

It can be seen, then, that Euler Circles afford a practicable way of testing the validity of certain uninvolved inferences. The technique is easy to teach, as was found in The Project for the Improvement of Thinking, which was carried on for two years in Evanston, New Trier, and Niles Township high schools in Illinois. In the selected geometry classes students were taught the technique of Euler Circles by using a set of exercises individually titled, *Classes and Class Membership, Class Inclusion and Exclusion and Ways of Illustrating These Relation-*

ships, How To Use Euler Circles To Test the Validity of Reasoning, Representing Other Statements (some p is q, some p is not q) *by Euler Circles,* and *Drawing Valid Conclusions.* The Euler Circle technique was chosen rather than the more elegant Venn diagram technique because it seems to be easier for students to visualize class inclusion and exclusion relationships.

FORMULAS OF VALID INFERENCE

We will now consider a topic which will enable us to understand the logic of the various strategies available to a student when he is trying to prove a proposition. Some of the strategies will be described in a subsequent section.

Let us suppose you are in a clothing store and hear this argument:

> If clothing material holds a press, this indicates that it is a synthetic fabric. And if the material wears well, this also indicates that it is a synthetic fabric. But worsted is not a synthetic fabric. So it won't hold a press and it won't wear well.

Is this a valid argument? Do the reasons prove the conclusion, i.e., worsted won't hold a press and it won't wear well, true? One way of testing the argument is to make an analogous argument using concepts which we may understand better than those in the preceding argument. Let us use geometric concepts. We shall make the following replacements:

'Clothing material holds a press' by 'two triangles are equiangular'

'It (clothing material) is a synthetic fabric' by 'they (two triangles) are similar'

'The (clothing) material wears well' by 'the two triangles are congruent'

'Worsted is not a synthetic fabric' by 'triangles ABC and DEF are not similar'.

With these substitutions, the argument becomes:

> If two triangles are equiangular, this indicates that they are similar. And if the two triangles are congruent, this also indicates that they are similar. But triangles ABC and DEF are not similar. So triangles ABC and DEF are not equiangular and they are not congruent.

Intuitively, this appears to be a valid argument. The reasons, if true, seem to prove the conclusion to be true. So, since this argument is analogous to the first argument, we feel justified in believing that the first argument is valid and the conclusion proved is a true statement.*

* One needs to remember that 'proof' and 'prove' are used here in the sense of mathematical or deductive proof; that is, the logical impossibility of the conclusion being false when the reasons are true.

Our justification is based on a principle of logic that validity depends only on the form of the argument and not on its subject matter. This means that two arguments of the same form are either both valid or both invalid regardless of their subject matter or content.

Now it is obvious that there are many arguments of many particular forms. If, then, we can identify those argument forms which are valid, any time we can fit a particular argument to a valid form we can be certain that it is valid and that its conclusion is true if the reasons given are true. Similarly, if we know certain invalid argument forms, we can be sure that any argument which fits one of these forms is invalid. It can be seen that this technique is powerful—by far the most powerful technique we have considered.

We can exhibit the form of quadratic equations like '$2x^2 + 3x + 1 = 0$' by '$ax^2 + bx + c = 0$' where 'a', 'b', and 'c' are parameters (variables). We call '$ax^2 + bx + c = 0$' a formula. Just as we can exhibit the form of certain algebraic sentences by replacing some of the numerals involved by variables, so we can exhibit the form of arguments by replacing the statements about a particular subject matter by variables. We shall

TABLE 3

Symbol	Possible Translations	Example
p		*Two triangles are equiangular.*
q		The triangles are similar.
$\sim p$	Not p. It is not the case that p. p is false	It is not the case that two triangles are equiangular.
$p \wedge q$	p and q	Two triangles are equiangular and the triangles are similar.
$p \vee q$	p or q*	Two triangles are equiangular or the triangles are similar.
$\sim(p \wedge \sim q)$	It is not the case that p and not q.	It is not the case that two triangles are equiangular and the triangles are not similar.
$p \rightarrow q$	If p then q. p only if q. p is a sufficient condition for q. q is a necessary condition for p	If two triangles are equiangular, then the triangles are *similar*.
$p \leftrightarrow q$	If p then q and if q then p. q if and only if p. p is equivalent to q. p is a sufficient and necessary condition for q.	Two triangles are equiangular if and only if the triangles are similar.

* The sense of 'or' is not the mutually exclusive sense, that is, p or q but not p and q, but the inclusive sense, namely, p or q and possibly p and q. To remove ambiguity the inclusive 'or' is sometimes written as 'and/or'.

call a sentence thus obtained a *formula*. The values (substitution instances) of the variables in such a formula will be statements. The values of the formula (as distinct from the values of the variables in the formula) will be arguments.

But before we can write such formulas and use them to test the validity of an argument, we will employ some symbols which will correspond to certain logical connectors in English, e.g., 'and', 'if ... then', 'or', and a symbol to correspond to the English 'not'. We shall use the dictionary as illustrated in Table 3 in which p and q are variables whose domain is statements.

We now can write some formulas for valid arguments (Table 4).

TABLE 4

Formula	Example
$[(p \to q) \wedge p] \to q$	If two triangles are equiangular, then they are similar and two triangles (i.e., ABC and DEF) are equiangular. Therefore, they (i.e., ABC and DEF) are similar.
$[(p \vee q) \wedge \sim p] \to q$	Lines A and B are either parallel or lines A and B intersect and lines A and B are not parallel. Therefore, lines A and B intersect.
$(p \to q) \to (\sim q \to \sim p)$	If two lines are perpendicular, then they form right angles. Therefore, if two lines do not form right angles, then they are not perpendicular.
$p \to \sim(\sim p)$	If a number is divisible by 10, then it is not the case that the number is not divisible by 10.

It now becomes apparent that if a teacher teaches his students the formulas most commonly used for proving statements, they will have a way of testing the validity of arguments. If an argument can be made to fit one of these formulas, it is valid. Most logic textbooks, especially those containing a chapter on symbolic logic, present many valid formulas. It can be seen, after a little thought, that the valid formulas are simply representations of laws of logic in symbols other than the words of ordinary language. We will see one advantage of this symbolism in the next section.

TRUTH VALUES AND TRUTH TABLES

The question, how do we know that certain formulas are valid formulas, naturally arises. Whatever answer is given must be based on the definition of 'valid inference', viz., 'a valid inference is one in which it is impossible for the reasons to be true and the conclusion false'. One answer is that for a valid formula we have never yet found a substitution instance in which the reasons were false and the conclusion true.

But this is an unsatisfactory answer for mathematics teachers. Who knows? Perhaps tomorrow someone may find just such an instance. A more satisfactory answer is that since we are concerned, according to the definition of 'valid argument', only with the truth and falsity of the reasons and conclusion, we *need* consider *only* these two as values of the variables in our formulas rather than an indefinite set of statements. Hence in testing a formula we will replace the variables by '*T*' or '*F*' which may be interpreted as 'true' and 'false' respectively. We can thus obtain all possible substitution instances whose reasons and conclusion have different truth values (i.e., true or false).

Before proceeding further, we shall have to establish some rules for determining the truth and falsity of compound sentences like $p \wedge q$, $p \vee q$, and $p \rightarrow q$—and even for a simple sentence like $\sim p$—when we know the truth values of p and q separately. We shall do this through the medium of ordinary English statements.

Let us take the statement, 'Angle A is a right angle'. According to the way we ordinarily speak, if 'Angle A is a right angle' is true, then we say that its contradictory, viz., 'It is not the case that Angle A is a right angle' is false. And if we know 'Angle A is a right angle' is false, then we say that its contradictory, viz., 'It is not the case that Angle A is a right angle' is true. In short, a statement and its contradictory have opposite truth values. We shall therefore use this, the way we ordinarily speak, *to define* the truth value of $\sim p$—the symbol for a negation—when we know the truth values of p. We shall say:

When p is	$\sim p$ is
T	F
F	T.

An array like this which enables us to ascertain the truth value of one statement variable when we know the truth value of another is called a *truth table*. This table satisfies the criteria for a function (see Chapter 3) so it is called a *truth-value function* or, more simply, a *truth function*. If p is considered to be the first member of the ordered pairs, the function provides a way of calculating the truth value of the second member, $\sim p$. On the other hand, if $\sim p$ is taken as the first member of the ordered pairs, we can calculate the truth value of the second member, p. The function thus obtained is the inverse of the first function.

Consider the statement, 'A measure of Angle A is 90°' along with the statement, 'Angle A is a right angle'. It is not possible for one to be true and the other false. Either both of these are true or both are false. This is the same as saying that they have the same *truth value*. State-

ments which have the same truth value are called *equivalent statements.* Let us symbolize two statements having the same truth value (being equivalent) as '$p \leftrightarrow q$'. We can use the following truth table to define $p \leftrightarrow q$:

When p is	and q is	$p \leftrightarrow q$ is
T	T	T
T	F	F
F	T	F
F	F	T.

We make use of this truth function in some proofs when we make substitutions. In any argument we may substitute for any statement an equivalent one without changing the validity of the argument. Examples of statements which are recognized as equivalent are the following:

This statement	is equivalent to	this statement
$3 + 4 = 7.$		$4 + 3 = 7.$
An angle may be copied by using a protractor or ruler and compass.		An angle may be copied by using ruler and compass or a protractor.
$x + 6 = 12$		$12 = x + 6$
$9x = 36$		$x = 4$
Lines m and n are parallel.		Lines m and n are not nonparallel.

In addition to the statement, 'Angle A is a right angle' let us take the statement, 'Angle B is an acute angle'. Each of these statements can be true or false (but neither can be both true and false). We can form a compound statement by joining these two statements by 'and', viz., 'Angle A is a right angle and Angle B is an acute angle'. We will call a statement like this formed by joining two or more statements by 'and' a *conjunction* and each of the component statements *conjuncts.*

What would we want to know about the truth of the conjuncts to be able to say, 'Angle A is a right angle and Angle B is an acute angle' is true? We are willing to say this conjunction is true if, but only if, each of its conjuncts is true. Otherwise, we say the conjunction is false. We shall now use these facts to define the truth values of $p \wedge q$—the symbol for a conjunction—when we know the truth values of p and q separately. Let us do this by means of the truth table (Table 5).

Notice that all possible combinations of the truth values of p and q are presented. The truth table is a calculus for arriving at the truth value of $p \wedge q$ given the truth values of p and q separately.

TABLE 5

When p is	And q is	$p \wedge q$ is
T	T	T
T	F	F
F	T	F
F	F	F

Let us now consider the compound statement, 'Angle A is a right angle *or* Angle B is an acute angle'. We shall call a statement like this a *disjunction*. A disjunction is a compound sentence consisting of two or more sentences joined by 'or'. Each of the component sentences of the disjunction will be called *disjuncts*.

What would we have to know about the truth of the disjuncts to be able to say that the disjunction, 'Angle A is a right angle or Angle B is an acute angle', is true? Unfortunately, we cannot say for sure, for 'or' is ambiguous. It is sometimes used in an inclusive sense so that 'Angle A is a right angle or Angle B is an acute angle' would be true if either of the disjuncts is true and also if both are true. 'Or' is also used in an exclusive sense as in 'Figure x is a triangle or (Figure x is) a quadrilateral'. This statement is true if either of the disjuncts is true, but is false if both are true. Which of these uses of 'or' shall we accept? We shall take the first because the inclusive sense will take care of part of the sense of the exclusive 'or', namely, that at least one of the disjuncts must be true for the disjunction to be true. This sense, partial though it is, is common to both uses of 'or'. We shall therefore *define* the truth values of $p \vee q$—the symbol for a disjunction—by the truth table (Table 6).

TABLE 6

When p is	And q is	$p \vee q$ is
T	T	T
T	F	T
F	T	T
F	F	F

It can be seen that the only case in which $p \vee q$ is false is when both p and q are false. Otherwise, it is true.

There remains one more kind of compound sentence to consider, represented by the sentence, 'If Angle A is a right angle, then Angle B is an acute angle'. We need a name for this kind of a sentence. Let us call it a *conditional*. A conditional is a compound sentence formed from two sentences by introducing one by 'if' and the other by 'then'. The

sentence following 'if' will be called the *antecedent* and the one following 'then' when 'then' is either written or implied, the *consequent*.

The conditional, 'If Angle A is a right angle, then Angle B is an acute angle', does not assert that the antecedent is true. Nor does it assert that the consequent is true. It asserts only that *if* the antecedent is true *then* the consequent is true. We would be willing to say that the conditional is true if both the antecedent and consequent are true. And we would be willing to say it is false if the antecedent is true and the consequent is false. Expressing this last statement differently, 'If Angle A is a right angle, then Angle B is an acute angle' is false when 'Angle A is a right angle and Angle B is not an acute angle' is true. Hence, 'If Angle A is a right angle, then Angle B is an acute angle' is true when 'Angle A is a right angle and Angle B is not an acute angle' is false.

We shall therefore use this meaning for stipulating the truth values of $p \rightarrow q$—the symbol for a conditional. Specifically, we shall define $p \rightarrow q$ as equivalent to $\sim(p \wedge \sim q)$. By this we mean that $p \rightarrow q$ shall have the same truth values as $\sim(p \wedge \sim q)$, i.e., when $\sim(p \wedge \sim q)$ is true $p \rightarrow q$ will be true and when $\sim(p \wedge \sim q)$ is false, $p \rightarrow q$ will be false. We shall now work out the truth table for $\sim(p \wedge \sim q)$ from the truth table for $p \wedge q$. By definition (see page 160) we see the results as illustrated in Table 5.

Now the truth value of $\sim q$ is the opposite of the truth value of q. So the truth table of $p \wedge \sim q$ is as illustrated in Table 7.

TABLE 7

When p is	And q is	$\sim q$ is	And $p \wedge \sim q$ is
T	T	F	F
T	F	T	T
F	T	F	F
F	F	T	F

In arriving at this truth table, we use the truth values of p and $\sim q$. The truth value of $p \wedge \sim q$ is 'true' only when both p and $\sim q$ are true and is 'false' otherwise.

The truth table of $\sim(p \wedge \sim q)$, which is the negation or contradictory of $p \wedge \sim q$, will be as illustrated in Table 8.

TABLE 8

When $p \wedge \sim q$ is	$\sim(p \wedge \sim q)$ is
F	T
T	F

Putting all these things together and remembering that $p \rightarrow q$ is defined as equivalent to $\sim(p \wedge \sim q)$, we have the results as illustrated in Table 9.

TABLE 9

When p is	And q is	$\sim(p \wedge \sim q)$ is	And so $p \rightarrow q$ is
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

From this table we can state the rules for determining the truth value of $p \rightarrow q$ when we know the truth values of p and q. $p \rightarrow q$ is true if and only if either p is false or q is true. Only when p is true and q is false is $p \rightarrow q$ false. We therefore see that $p \rightarrow q$ is also equivalent to $(\sim p) \vee q$.

TESTING THE VALIDITY OF FORMULAS

We are *now ready to test inference* formulas. Let us begin with one we commonly use: $[(p \rightarrow q) \wedge p] \rightarrow q$. An instance of this is: if $x = 10$, then $2x = 20$, and $x = 10$. Therefore $2x = 20$. We shall set up a truth table for this, beginning with all possible combinations of the truth values of p and q in Columns 1 and 2 (Table 10). Next we fill in Column

TABLE 10

p	q	$p \rightarrow q$	$(p \rightarrow q) \wedge p$	$[(p \rightarrow q) \wedge p] \rightarrow q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

3 by using the rule that $p \rightarrow q$ is true if either p is false or q is true. We then fill in *Column* 4 using the truth values in Columns 3 and 1 and the rule that the truth value of a conjunction is true if both conjuncts are true; otherwise the truth value is false. Finally, we *fill* in the truth values in Column 5 using those in Columns 4 and 2. We find that truth value of this formula is always 'T' or true. This means that all the substitution instances (arguments) of this formula are valid. In general a formula is valid if and only if the truth value of the formula is al-

ways 'T', for obviously it is not then possible for the premises to be true and the conclusion false.

Let us test the validity of the formula, $[(p \vee q) \wedge p] \rightarrow \sim q$.

In Columns 1 and 2 of Table 11, we form all possible combinations of

TABLE 11

p	q	$\sim q$	$p \vee q$	$(p \vee q) \wedge p$	$[(p \vee q) \wedge p] \rightarrow \sim q$
T	T	F	T	T	F
T	F	T	T	T	T
F	T	F	T	F	T
F	F	T	F	F	T

truth values of p and q. Using Column 2 and the rule that the truth value of $\sim q$ is the opposite of the truth value of q, we fill in Column 3. Using Columns 1 and 2 and the rule that $p \vee q$ is true if either p or q is true, then, we fill in Column 4. Using Columns 4 and 1 and the appropriate rule, we fill in Column 5. Using this column and Column 3 and the appropriate rule, we fill in Column 6. Since the truth value in Column 6 is not always 'T', this proves that the formula, $[(p \vee q) \wedge p] \rightarrow \sim q$ is invalid.

As exercises, the reader may test the validity of the following formulas.

$$[(p \rightarrow q) \wedge \sim p] \rightarrow \sim q \qquad \text{(invalid)} \qquad (1)$$

If we assert the falsity of the antecedent of a true conditional, we cannot infer the falsity of the consequent.

As an example of this invalid formula, we have:

If rectangles have four right angles, squares have four right angles.
Rectangles do not have four right angles.
Therefore squares do not have four right angles.

The first statement, which is conditional, is true. The second statement asserts the falsity of the antecedent of the conditional. But we cannot validly infer the falsity of the consequent, viz., squares do not have four right angles.

$$[(p \rightarrow q) \wedge q] \rightarrow p \qquad \text{(invalid)} \qquad (2)$$

If we assert the truth of the consequent of a true conditional, we cannot infer the truth of the antecedent.

As an example of this invalid formula, consider:

If arithmetic is easy for Dorothy, algebra is easy for Dorothy.
Algebra is easy for Dorothy.
Therefore arithmetic is easy for Dorothy.

The first statement is the conditional. The second statement asserts the truth of the consequent of the conditional. Let's say both of the statements are true. Even so, a conclusion which asserts the truth of the antecedent, viz., arithmetic is easy for Dorothy, does not necessarily follow.

$$[(p \rightarrow q) \wedge \sim q)] \rightarrow \sim p \qquad \text{(valid)} \qquad (3)$$

If we assert the falsity of the consequent of a true conditional, we can infer the falsity of the antecedent.

An example of this formula is the following:

For all values of a and b,
if a is greater than b, b is less than a.
b is not less than a.
Therefore, a is not greater than b.

The second statement asserts the falsity of the consequent of the true conditional. The conclusion, which is the assertion of the falsity of the antecedent, follows necessarily.

$$(p \rightarrow q) \rightarrow (q \rightarrow p) \qquad \text{(invalid)} \qquad (4)$$

The converse of a true conditional (i.e., the proposition formed by interchanging the antecedent and consequent) is not necessarily true.

It is true that if a number is an even number, it is a whole number. But the converse, viz., if a number is a whole number it is an even number, is not true. It is true that if a number is an even number it is divisible exactly by 2. The converse, viz., if a number is exactly divisible by 2 it is an even number, is also true. These examples show that the converse of a true conditional may be true and may not be true. We cannot infer validly one way or the other.

$$(p \rightarrow q) \rightarrow (\sim p \rightarrow \sim q) \qquad \text{(invalid)} \qquad (5)$$

The inverse of a true conditional (i.e., the proposition formed by negating the antecedent and consequent) is not necessarily true.

It is true that if a teacher does not know mathematics, he will be a poor mathematics teacher. But the inverse of this conditional, viz., if a teacher does know mathematics he will not be a poor teacher, is not always true.

$$(p \rightarrow q) \rightarrow (\sim q \rightarrow \sim p) \qquad \text{(valid)} \qquad (6)$$

The truth of the contrapositive of a proposition (i.e., the proposition formed by negating the antecedent and consequent and interchanging them) follows necessarily from the truth of the proposition.

This expresses the same idea as Formula 3 and both are equivalent to the formula $(p \to q) \leftrightarrow (\sim q \to \sim p)$.

We know that if a triangle is equilateral, it is isosceles. We can validly conclude that if a triangle is not isosceles, it is not equilateral. The second statement is the contrapositive of the first.

$$(p \wedge \sim p) \to q \qquad \text{(valid)} \tag{7}$$

This formula indicates why inconsistencies are to be avoided in any deductive system like mathematics. As the formula shows, if inconsistencies are allowed, any proposition whatsoever can be proved. If $1 = 1$ and $1 \neq 1$, we can readily prove that you own the Empire State Building.

In summary, the truth table technique can be used to test the validity of formulas whose substitution instances are arguments. Once valid formulas have been identified, these can be used to test the validity of particular arguments. If a particular argument fits a valid formula it is valid. If a particular argument has a form whose validity is unknown, a formula for the argument can be written and its validity tested by forming its truth table. For example, by writing the formula $[(p \to q) \wedge (r \to q) \wedge \sim q] \to (\sim p \wedge \sim r)$ for the argument on page 156 and testing the formula, we find that it is a valid formula. Therefore, the argument which is an instance of this formula is valid.

How much about formulas of inference and truth tables should be taught students in high school mathematics courses remains to be decided. Hendrix[23] reports teaching such topics to an eleventh grade class in solid geometry. A report[24] written by the students in an algebra II class indicates that such topics are within the comprehension of eleventh and twelfth grade students. There are appearing on the market textbooks[25] intended for senior high school and junior college which present these topics. Bright students in mathematics will enjoy studying these topics.

STRATEGIES OF PROOF

When a teacher or student wants to prove a proposition, he will find it useful to know various plans or strategies of proof. These can be derived from some of the formulas of inference already discussed.

Counterexample. As pointed out earlier, finding a counterexample is a stratagem which can be used to disprove a proposition. For example, if one student argues that any triangle whose sides are in the ratio 3:4:5 has angles of 30°, 60°, and 90°, another student can disprove this statement by exhibiting a 3-, 4-, 5-triangle and either measuring the angles or computing their values by trigonometry. Suppose someone asserts

that if a student studies every day, he will pass all his tests. This proposition can be disproved by citing one student who studied every day and did not pass all his tests. More generally, to disprove a proposition, either prove its contradictory or prove that its acceptance leads to a contradiction of a postulate or theorem.

Detaching an Antecedent. A very common form of argument is illustrated by the following: If two numbers are added, the order in *which they are* added *makes no difference. (The* numbers) *7 and 4* are added. Therefore, the order in which (the numbers) 7 and 4 are added makes no difference. This is a valid argument because it fits the valid inference formula, $[(p \rightarrow q) \land p] \rightarrow q$. Since the two reasons of this valid argument are true, the truth of the conclusion is proved. You can verify this by examining the first line of truth values in Table 10 on page 163.

This suggests a stratagem for proving a proposition, q. Try to find *a true conditional in which q is the consequent, e.g., if p, then q.* Then try to prove that the antecedent, p, is true. If p is true, you can assert the truth of the consequent, q. This stratagem can also be employed when the proposition, q, to be proved is a compound proposition like $r \land s$, $r \lor s$, or $r \rightarrow s$.

Detaching an antecedent is also spoken of as *modus ponens.*

Developing a Chain of Propositions. Sometimes we can discover how to prove a proposition, p, by realizing that we can prove p if we can prove q. And we can prove q if we can prove r. For some reason we know that r is true. We then can make the proof. Its validity is justified by the valid inference formula:

$$[r \land (r \rightarrow q) \land (q \rightarrow p)] \rightarrow p.$$

A moment's reflection reveals that this stratagem is a repeated application of the stratagem of detaching the antecedent. r being true and $r \rightarrow q$ allows us to detach r and assert q. q and $q \rightarrow$ p allows us to detach q and assert that p is true.

Proving a Conditional. To prove the proposition, if a triangle is isosceles, the angles opposite the equal sides are equal, the proof proceeds by assuming the antecedent and then arguing that the consequent is true. The validity of the parts of the argument or the argument as a whole can be checked by the procedure described in the previous section. If the argument is successful, this proves the proposition. In general, to prove a conditional, p $\rightarrow q$, assume the antecedent, p, and argue that the consequent, q, is true.

A second stratagem for proving a conditional, p $\rightarrow q$, is to prove its

contrapositive, viz., $\sim q \rightarrow \sim p$. This stratagem is justified by the valid inference formula, $[(\sim q) \rightarrow \sim p] \rightarrow (p \rightarrow q)$. Suppose we wish to prove the proposition, if two lines are cut by a transversal making the alternate interior angles equal, the lines are parallel. We write its contrapositive, i.e., the proposition formed by interchanging the antecedent and the consequent and negating each. In the case of the proposition we are considering, the contrapositive may be stated as: if two lines are not parallel, then the alternate interior angles formed by a transversal are not equal. We then prove this proposition by using the stratagem described in the preceding paragraph. This enables us to argue that the original proposition, viz., if two lines are cut by a transversal making the alternate interior angles equal, the lines are parallel, is true. For our reason we cite the law of contraposition, viz., $(p \rightarrow q) \leftrightarrow [(\sim q) \rightarrow (\sim p)]$.

Sometimes the consequent of the conditional is a conjunction as in the proposition, if the bisector of the angle of an isosceles triangle which is not one of the equal angles intersects the side opposite this angle, it is perpendicular to that side and bisects the side. A stratagem which can be used is repeated application of the stratagem for proving a simple conditional. Assume the antecedent, and argue that one of the conjuncts of the consequent is true. Then repeat the procedure for each of the other conjuncts.

A third way of proving a conditional of the form $p \rightarrow q$ is to find some statement r such that you can prove $p \rightarrow r$ and $r \rightarrow q$ perhaps by using either of the two strategies just described. You can then infer $p \rightarrow q$, justifying the inference by the valid formula:

$$[(p \rightarrow r) \wedge (r \rightarrow q)] \rightarrow (p \rightarrow q) .$$

You may want to assure yourself that this formula is valid by forming its truth table. If you do, you will have eight rows in the table.

The advice "clear of fractions" which many algebra teachers give students who are considering how to solve an equation like $1/3 + 1/4 = 1/x$ illustrates the use of this stratagem. If we want to prove that for every value of x if $1/3 + 1/4 = 1/x$ then $x = 12/7$, we first prove that if $1/3 + 1/4 = 1/x$, then $4x + 3x = 12$. Next we prove that if $4x + 3x = 12$, then $x = 12/7$. Hence, we can infer validly that if $1/3 + 1/4 = 1/x$, then $x = 12/7$.

Reductio ad Absurdum. This stratagem is not hard, but many mathematics students do not understand it and think there is something unsound about it. Yet it is powerful and is frequently used. The stratagem consists in accepting the contradictory of the proposition to be proved and proving that this leads to an inconsistent proposition, e.g.,

$q \wedge \sim q$, which is necessarily false. The reason for refusing to admit such propositions was explained on page 166.

The inference formula justifying this kind of indirect proof is interesting. Symbolizing the proposition we wish to prove as p, its contradictory becomes $\sim p$. We *prove that* $\sim p$ *is false* which means that p is true. The formula is:

$$\sim(\sim p) \rightarrow p.$$

Since the stratagem leads to an inconsistent statement, e.g., $q \wedge \sim q$ this method is known as *reductio ad absurdum*—reduced to an absurdity.

Indirect Proof. Suppose a girl has only three dresses, say A, B, and C, which are appropriate to wear to a dance, and she is undecided on which to select. The color of dress A will clash with the color of the corsage her date is sending so this dress is eliminated. Dress B has a loose hem, and the girl does not have time to sew the hem. So this dress is eliminated. Dress C, the only one left, is the one she selects. This case illustrates indirect reasoning. From it we can form a stratagem of proof. First, formulate a true disjunction whose disjuncts are exhaustive of the domain under consideration. An example is 'Lines m and n are parallel, or lines m and n intersect, or lines m and n are skew'. One of these disjuncts has to be true. Next, prove false one by one all the disjuncts except one. Often this is done by *reductio ad absurdum*. Finally, use the valid inference formula:

$$[(p \vee q \vee \cdots \vee t) \wedge ((\sim q) \wedge \cdots \wedge (\sim t))] \rightarrow p$$

and thereby prove the remaining disjunct true.

Proving a Statement of Equivalence. Suppose we wish to prove that $2x + 6 = x - 7$ and $x = -13$ are equivalent equations, i.e., $(2x + 6 = x - 7) \leftrightarrow (x = -13)$. By this we mean that for every value of x, $2x + 6 = x - 7$ and $x = -13$ have the same truth value, i.e., both true or both false. We remember (see page 160) that a statement of equivalence is a statement of double implication, that is, '$p \leftrightarrow q$' means '$p \rightarrow q$ and $q \rightarrow p$'. In the case of our example, we can prove the two equations equivalent if we can prove: (1) if $2x + 6 = x - 7$ then $x = -13$ and (2) if $x = -13$, then $2x + 6 = x - 7$. Both of these are conditionals, hence whatever strategies are applicable to proving conditionals are applicable to proving statements of equivalence.

Mathematical Induction. Like indirect proof, mathematical induction is a stratagem used by students but often not understood by them. Some students are able to reproduce a proof of, say, the theorem $1 + 2 + \cdots + n = n(n + 1)/2$ by following the rules given in the

textbook and imitating the sample problem. But they seldom understand the logic of this kind of proof.

Let us introduce the logic of this stratagem by an analogy. Suppose there is a row of dominoes. Each domino is standing on end and is close enough to the one on its left so that if it is knocked over it will knock over the one on its left. The row extends as far as the eye can see. Let us assume someone designates a particular domino in the line which is far from us and tells us to knock it down. We can do this simply by knocking the first domino down, for we know that if any domino is knocked down it will knock down the one on its left. The analogy with mathematical induction can now be seen. Take the theorem stated in the previous paragraph, viz., $1 + 2 + \cdots + n = n(n + 1)/2$. Proving that if the theorem holds for any integer, it holds for the next integer can be compared with arranging the dominoes so that if one is knocked over the next one will also be knocked over. Providing that the theorem holds for 1 can be compared with knocking the first domino down.

It has been mentioned earlier in this chapter that mathematical induction does not result in a probable inference but rather in a necessary inference. To put it differently, mathematical induction is not induction, but rather deduction. As will be seen shortly, the stratagem of mathematical induction can be subsumed under the stratagem of detaching the antecedent.

Courant and Robbins[24] say:

> Let us suppose that we wish to establish a whole infinite sequence of mathematical propositions
>
> $$A_1, A_2, A_3, \cdots$$
>
> which together constitute the general proposition A. Suppose that a) by some mathematical argument it is shown that if r is any integer and *if* the assertion A_r is known to be true, then the truth of the assertion A_{r+1} will follow, and that b) the first proposition A_1 *is* known to be true. Then all the propositions of the sequence must be true, and A is proved.

It can be seen that an argument by mathematical induction consists of a conditional whose antecedent, to use Courant and Robbins' notation, is the conjunction of A_1 and $A_r \rightarrow A_{r+1}$ and whose consequent is A. Each of the conjuncts is proven true. We then detach the antecedent and assert the consequent, A.

THE RELATIONS BETWEEN TRUTH OF THE STATEMENTS CONSTITUTING THE INFERENCE AND THE VALIDITY OF THE INFERENCE

Once students have acquired the concept of validity of an inference and truth of a statement, they can be taught the relations between the

truth of the statements constituting the inference and the validity of the inference. We may identify four relations:

1. A person may know in some way that the reasons in an argument are all true. He also may know that the inference is valid. He then is justified in stating that the conclusion is true.

2. A person may know that an inference is valid. He also may know in some way that the conclusion is false. He then is justified in stating *that at least one of the reasons is false.*

3. A person may know in some way that all the reasons in an argument are true, but that the conclusion is false. He then is justified in stating that the inference is invalid.

4. A person may know in some way that the conclusion of a proof is true. He may also know that the inference is valid. But *from only this knowledge* he cannot know whether the reasons are true or not.

These relationships follow from the definition of 'valid inference'. The first relation is the one we use to pass from knowledge we presently have to new knowledge. We use it both inside and outside the classroom. Most of the arguments we use and those used on us are enthymemes, that is, arguments in which one or more reasons are unstated, or the conclusion itself is unstated. But the reasons or the conclusion usually can be stated by the maker of the argument if requested.

A common misconception related to this first relation should be mentioned. Many novices believe that if all the reasons are false and the inference is valid, the conclusion is necessarily false. This belief may be explained as erroneously assuming the truth of the inverse of the first relation when stated as, 'If all the reasons are true and the inference is valid, the conclusion is true'. That this inverse is not necessarily true may be demonstrated by the simple syllogism: All squares are circles. All circles are quadrilaterals. Therefore, all squares are quadrilaterals. Both reasons are false. The inference is valid and the conclusion is true.

The second relation may be stated as, 'If an inference is valid and the conclusion is false, then at least one of the reasons is false'. It is used when we doubt the truth of a certain statement but the person advancing it does not. We are careful to see that his inference based on this statement is valid and hope to show that the necessary conclusion is materially false. We then charge that the statement at issue is false, providing, of course, we and he agree that the other reasons are all true. The expression "All right, let's assume that this statement is true. Now let's see what follows" or similar expressions are often cues that this relation between the truth of the constitutive statements of an inference and the validity of the inference is being considered.

PROOF AND A LOGICAL SYSTEM

The capstone of an understanding of proof is an understanding of the nature of a logical system like mathematics and the role proof plays in developing this system. Allendoerfer implies this when he says[27] "At some stage in the high school mathematics curriculum there should be a serious discussion of deductive systems *per se*, and later applications of this to mathematics and to nonmathematical situations should be used to reinforce the understanding of the students about deductive methods."

Let us consider briefly the nature of a deductive system which might be a part of mathematics. The symbols of such a system can be grouped in two sets. The smallest set consists of the primitive symbols, e.g., 'point', 'join', 'set', 'operator', and 'x'. These are undefined and of necessity must remain undefined within the system if circularity is to be avoided.

The second set of symbols consists of those symbols introduced into the system by definitions. The defining part of the definition contains only primitive symbols or symbols previously defined in terms of the primitives. The definitions may be regarded as stipulations as to how the new symbols are to be used. For example, we might want to use a simpler term (symbol) than 'a rectangular parallelepiped all of whose edges are equal'. If so, we could define 'cube' by saying that wherever 'a rectangular parallelepiped all of whose edges are equal' is used, the term 'cube' may be used to replace it and *vice versa*. Or, more formally, we could write, "cube $=$ *df* a rectangular parallelepiped all of whose edges are equal." The symbol '*df*' indicates that the sentence is a definition, that the term following the '*df*' is the defining part of the definition, and the term on the other side of the '$=$' is the term defined.

The definitions may be regarded as the rules of syntax of the system. The analogy between these rules and the rules of syntax of English is interesting. To the set of primitive symbols of the system, there corresponds the set of letters of the alphabet and the punctuation marks. To the rules of formation of new symbols, there correspond the spelling rules and the rules of grammar. The definitions (rules of syntax) become the criteria for a well-formed expression. They provide an effective test for determining whether or not a given expression is a part of the system. For example, '$3/0$' and '$\rightarrow\ =\ \sim$' might not be admitted to the system, for one could point to no rule defining these symbols.

The sentences of the system may also be grouped in two sets. In the smallest set are those which are the primitive sentences—those from which all the other sentences in the system are deduced. These are called the *axioms* (or sometimes are called *postulates*) of the system.

The second set of sentences are those which are deduced from the axioms by using agreed-upon rules of inference. They are called *theorems*. Proof of the first theorem is a matter of deducing a conditional in which the antecedent is a conjunction of some of the axioms, and the consequent is the theorem. In the proof of subsequent theorems, the antecedent of the conditional is a conjunction of axioms or theorems previously proven and the consequent is the particular theorem to be proved.

A necessary condition for a satisfactory set of axioms is that they do not lead to inconsistent (contradictory) statements. The reason for demanding this condition has been stated on page 166. A further condition which is not necessary (unless one first demands elegance) but is desirable is the independence of the axioms. That is, none of the axioms should be able to be proved from some of the others.

The deductive system which has been described is an uninterpreted system. This is because no meaning, in the semantic sense, was given the primitive symbols. If we want to interpret it, we must state rules of correspondence which relate the primitive symbols to "things" outside the system. For example, we might decide that 'x' is to correspond to any point on paper made by a pencil and 'A' is to correspond to any line on paper made by a pencil. Then the axioms:

1. There exist at least two x's.
2. Every A is a collection of x's.
3. If x_1 and x_2 are x's, then there is one and only one A containing x_1 and x_2.

would be interpreted as:

1. There exist at least two points made by a pencil on paper.
2. Every line on paper made by a pencil is a collection of points made by a pencil on the paper.
3. If there are two points made by a pencil on paper, then one and only one line on the paper made by a pencil contains these two points.

In such a deductive system, the role of proof is that of providing a test of whether or not a given proposition has the same truth value as that of the axioms. If we arbitrarily assign the truth value 'true' to the axioms, we can assign the same truth value to any proposition resulting from the use of the axioms and the rules of inference agreed upon. Hence, from another point of view, proof is a way of testing whether or not a given proposition, other than an axiom, is part of the deductive system.

This is a brief description of the nature of a logical system like mathematics. It is overly simplified because the purpose of this section of the chapter is to suggest ways of giving students an understanding of the nature of a mathematical system rather than to explicate the concept. For a more complete explication, Allendoerfer's chapter in the twenty-

third yearbook of the National Council of Teachers of Mathematics is recommended[27].

The idea of proof is probably easier for students to understand than is the idea of structure. As has been stated earlier, the concept of proof begins to form when the child learns to ask why and to accept some reasons and reject others. But the idea of structure is a sophisticated idea which probably will not mature until students have studied considerable mathematics in college. Nevertheless, a start can be made in high school in both algebra and geometry. In the past where such a start has been made, it usually has been made in demonstrative geometry. But whether pedagogically this is best remains to be determined by experimentation. Euclidean geometry has the advantage of not having so much subject matter that is prerequisite to the study of subsequent mathematics as algebra. Hence, the teacher does not feel he is handicapping his students in future mathematics courses if he uses class time to teach them an understanding of a mathematical system. But there is much to be said for beginning the study of the structure of mathematics in the first course in algebra. This course essentially is a study of the properties of a field (see Chapter 2, "Number and Operation"). It should be possible to consider rational numbers as a set of numbers which satisfies the axioms defining a field. Then when the need for irrational numbers is demonstrated, these together with the rational numbers can be shown to constitute another field, the field of real numbers.

There are rather simple systems which can be developed in a course in mathematics in the twelfth grade and which can serve admirably to illustrate mathematical structure. Ideas for doing this can be obtained from Chapters III and V in the twenty-third yearbook of the National Council of Teachers of Mathematics[21] cited above. Deliberate teaching of the nature of a mathematical system is increasingly being incorporated in textbooks intended for use in first-year college mathematics courses.[25]

Let us turn to reported attempts by high school teachers to teach students about the nature of a mathematical system. One example is that by Fawcett.[29] Each of Fawcett's students built his own system of geometry, which he called his "theory of space," without use of a textbook. The year's work began with discussion of nonmathematical topics aiming at showing the students the need for clear definition of key words. Students were led to see the effect of definitions on conclusions; how a change in a definition could lead to a different conclusion. They also became aware of the ambiguity, vagueness, and emotional overtones associated with many terms in our ordinary language. Fawcett then guided them into a realm of ideas—ideas about space—in which these

three difficulties of language would be minimized. Again, by careful developmental teaching, the students were led to discover the unavoidability of undefined terms in choosing names for ideas about space. Each student then chose a set of terms he intended to use as undefined in his theory of space and defined other terms by means of these and certain nontechnical words in the English language. He also selected a set of *propositions which he took as his basic assumptions (axioms)*. He then proceeded to formulate propositions that appeared *true and to prove* those that could be proved. These then became the *theorems of his* unique theory of space or geometry.

Fawcett summarized his general procedures as follows:

1. No general text was used. Each pupil developed his own text and was given the opportunity to develop it in his own way.
2. The undefined terms were selected by the pupils.
3. No attempt was made to reduce the number of undefined terms to a minimum.
4. The terms needing definition were selected by the pupils and the definitions were an outgrowth of the work rather than the basis for it
5. Definitions were made by the pupils. Loose and ambiguous statements were refined and improved by criticisms and suggestions until they were accepted by all pupils.
6. Propositions which seemed obvious to the pupils were accepted as assumptions.
7. These assumptions were made by the pupils and were recognized by them as the product of their own thinking.
8. No attempt was made to reduce the number of assumptions to a minimum.
9. The *detection of implicit or tacit* assumptions was encouraged and recognized as important.
10. No statement of anything to be proved is given the pupil. Certain properties of a figure are assumed and the pupil is encouraged to discover the implications of these assumed properties.
11. No generalized statement is made before the pupil has had an opportunity to think about the implications of the particular properties assumed. The generalization is made by the pupil after he has himself discovered it.
12. Through the assumptions made the attention of all pupils is directed toward the discovery of a few theorems which seem important to the teacher.
13. Assumptions leading to theorems that are relatively unimportant are suggested in mimeographed material which is available to all pupils but not required of any.
14. Matters of common concern such as the selection of undefined terms, the *making of definitions*, the *statement of assumptions* and the generalizing of an implication are topics for general discussion while periods of supervised study provide for individual guidance.
15. The major emphasis is not *on the theorems proved* but rather on the method of proof. This method is generalized and applied to non-mathematical situations.

Of particular significance, as one will see if he reads Fawcett's report, was Fawcett's demonstration by repeated well-conceived exercises about nonmathematical topics that the kinds of reasoning pervading mathematics pervade nonmathematical topics also. Students saw the effect of definitions in insurance, sports, tax laws, and school problems. They ferreted out hidden assumptions in reasoning employed in advertising, editorials, articles in periodicals, lectures, and political statements. The purpose of this kind of teaching was to enable the students to transfer more readily their knowledge of proof to new contexts. That such transfer occurred and that the students did acquire an understanding of the nature of a mathematical system is evidenced by the protocols Fawcett presents in the chapter, "Evaluation." This book contains a wealth of ideas for a teacher who has interests similar to those of Fawcett.

Should a geometry teacher not want to embark on such a free-wheeling approach as Fawcett's he still can give his students an experience in building a system of geometry for themselves. One article[33] describes a possible approach within the conventional geometry course whose scope and sequence are determined by a textbook. When the chapter on parallel lines is reached, the teacher can abandon the textbook temporarily. He can draw two parallel lines on the blackboard and a transversal. Using letters to identify the angles formed, he can get the students to state propositions which appear to be true at first sight. These may be written on the board. Drawing a perpendicular to one of the parallels from the point of intersection of the transversal and the other perpendicular will uncover other propositions. So will drawing a perpendicular to one of the parallels from the midpoint of the segment of the transversal included between the two parallels and extending the perpendicular until it meets the other parallel.

The students can then be told to select a small number of these propositions as their "axioms," and to try to discover and prove propositions about parallel lines and geometric figures composed in part of parallel line segments by using these axioms and whatever definitions they decide to introduce into their system. Since it is unlikely that even two students will select the same set of axioms and defined terms, each student's mathematical system will be unique. Each student will see how a mathematical system is developed by developing one himself.

Another occasion when this system-building can be used is in the study of circles. The same procedure described above can be used to get the work under way. Again each student probably would develop a unique mathematical system identifying many theoretically useful

concepts, e.g., those to which the names 'chord', 'central angle', 'tangent', 'secant', and 'inscribed angle' are ordinarily assigned.

A provocative variation of the geometry of the circle could be made by the teacher telling the students to consider the surface enclosed by the circle as the plane and to build a system of geometry descriptive of this plane. Such a system would force reinterpretation of such terms as 'plane' and 'line'. Both a plane and a line would be finite in extent. Of course, rather than reinterpret the terms, new names could be given the new ideas. But if the old terms are reinterpreted, this would foster an understanding of a mathematical system, for the students would see that undefined terms like 'plane' and 'line' can be interpreted in more than one way. By direction of the students' thinking, the teacher could initiate an understanding of the ideas of consistency, independence, and categoricalness of a set of axioms.

An experiment similar to Fawcett's but not as extensive was carried out by Henderson in a one-semester course in solid geometry.[10] The part of this experiment relevant to this discussion was the two experiences the students had in building their own systems of geometry. When the chapter on prisms was reached, each student put away his textbook and did not refer to it as far as the teacher could determine for the two weeks spent on the unit. The teacher brought to class models of several different prisms. By encouraging the students to state the relationships that appeared to be true about the prisms a set of propositions was developed. The teacher then suggested that each student select a group of these for the axioms of his system of prismatic geometry and try to prove some theorems.

Forty-three different theorems were proved by one or more of the sixty-three students in the classes. The smallest number of theorems proved by a student was four; the largest was twenty-four. The median was ten. Class time was spent either in supervised study or evaluating the proofs of propositions that students had developed. Indirect proof was often used. Terms were defined. If a concept was invented by one student which seemed fruitful, it was adopted by other students. An example of this was concept named by 'perpenprism' which the inventor defined as we would define 'right prism'. Many of the concepts ordinarily used in the theorems about prisms were discovered by the students although the ideas were usually named differently. In this naming the use of metaphors was interesting, e.g., 'box' for 'rectangular parallelepiped' and 'dispersoid' and 'wedge' for 'triangular prism'.

Common errors in proof were assuming the converse or inverse of a

proposition, reasoning in a circle, being unaware of hidden gratuitous assumptions, and basing the proof of a general theorem on a special case. As these errors were identified, they were discussed and the principle of logic involved stated. Students who made the errors corrected them.

At the end of the two-week experience, the solid geometry text was brought out, and, by leafing through the chapter on prisms, the students saw the conventional names and definitions for the concepts they had identified. In this reading of the chapter some students saw theorems in the book which they had discovered and proved in their own systems. This was a source of considerable satisfaction to them. One boy in all sincerity thought that either the authors were unaware of a particular theorem he proved or failed to realize its importance since they had not included it in the text.

When the study of pyramids was reached, another experience in building a mathematical system was provided. The procedure was similar to that followed in the construction of the *prismatic geometry* with the content being pyramids instead of prisms. It was gratifying to notice the greater sophistication with which the students approached their second building of a mathematical system.

The experiments reported by Fawcett and others have the advantage of teaching students about the nature of a logical system by firsthand experience in building such a system. The students learn in a dramatic way the unavoidability of undefined terms and unproved propositions, the freedom to name and define, and the significance of theorems as consequences of the axioms. They learn conformance to rules of logic as the criterion of correct proof rather than intuition. Excursions like these do not necessarily result in a loss of knowledge of the theorems of geometry. They do result in a concept of a logical system and the role proof plays in developing the structure.

SUMMARY

The idea of proof is one of the pivotal ideas in mathematics. It enables us to test the implication of ideas thus establishing the relationship of the ideas and leading to the discovery of new knowledge.

To the unsophisticated, 'proof' is practically synonymous with 'what convinces me'. A statement is "proved" when such a person is convinced that it is true. Such a concept makes proof a subjective and personal matter. The purpose of considering proof as a major idea in mathematics is to lead students from such a subjective concept to a more objective one, a concept based on criteria interpersonal in nature.

People are continually called upon to make inferences. To do this they pass either consciously or unconsciously from one or more propositions, called *evidence* or *reasons*, to another proposition called a *conclusion*. If they rationally reconstruct the reasoning they or someone else employs, two questions always are central: (1) Are the reasons in fact true, and (2) Does the conclusion necessarily follow? It was toward these two questions that the chapter was directed. These were used as the bases of organizing the discussion because of their importance and because it was felt all teachers from the primary grades through high school can make a contribution toward an understanding of them.

The material truth of a statement, if it has such a property, is established by inferences which are probable. The necessity of a conclusion is established by laws of logic. In developing these ideas it is a matter of the teacher's beginning with the unverbalized and inconsistent ways by means of which the child says, in effect, "It is true that . . ." or "Therefore, . . ." In the case of probable inference, the starting point is personal experience—very convincing but unreliable—and authoritative opinion. In the case of necessary inference, it is the sentences containing 'why' and 'because' that the child makes. As he progresses through the grades, he is taught to check his judgments. He is taught to check probable inferences by securing data from experience, and to check necessary inferences by applying principles of logic. Armed with this knowledge, the student should be a more disciplined thinker whether he is thinking about mathematics or everyday problems.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in this chapter.

NOTES

1. Keyser, Cassius J. *Thinking about Thinking.* New York: E. P. Dutton & Co., 1926.
2. Whitehead, Alfred North. *The Function of Reason.* Princeton, N. J.: Princeton University Press, 1929. p. 8.
3. National Council of Teachers of Mathematics. *The Place of Mathematics in Secondary Education.* Fifteenth Yearbook. New York: Bureau of Publications, Teachers College, Columbia University, 1940. p. 24.
4. Gesell, Arnold, and Ilg, Frances L. *The Child from Five to Ten.* New York: Harper and Brothers, 1946. p. 447.
5. Piaget, Jean. *The Psychology of Intelligence.* Translated by Malcolm Piercy and D. E. Berlyne. New York: Harcourt, Brace & Company, 1950 p. 123.
6. Searles, Herbert L. *Logic and Scientific Methods.* New York: The Ronald Press Company, 1948. p. 163.
7. Schaaf, Oscar Frederick. *Student Discovery of Algebraic Principles as a Means of Developing Ability to Generalize.* Unpublished Doctoral Dissertation. Columbus: Ohio State University, 1954 p. 17.

8. Polya, G. "Induction and Analogy in Mathematics." *Mathematics and Plausible Reasoning*. Princeton, N. J.: Princeton University Press, 1954. Vol. 1, p. 17.
9. Polya, G. "Induction and Analogy in Mathematics." *Mathematics and Plausible Reasoning*. Princeton, N. J.: Princeton University Press, 1954. Vol. 1, p. 17–21.
10. Board of Regents of the Smithsonian Institution. *Annual Report of the Board of Regents of the Smithsonian Institution, 1909.* Washington, D. C.: Government Printing Office, 1919. p. 253–60.
11. Abbott, Edwin. *Flatland*. New York: Dover Publications, 1952. p. 85–95. Kasner, Edward, and Newman, James. *Mathematics and the Imagination*. New York: Simon & Schuster, 1940. p. 124–35.
12. Wertheimer, Max. *Productive Thinking*. New York: Harper and Brothers, 1945. p. 164.
13. Polya, G. "Induction and Analogy in Mathematics." *Mathematics and Plausible Reasoning*. Princeton, N. J.: Princeton University Press, 1954. Vol. 1, p. v.
14 Polya, G. "Induction and Analogy in Mathematics." *Mathematics and Plausible Reasoning*. Princeton, N. J.: Princeton University Press, 1954. Vol. 1, p. v–vi.
15. Segri, Emilio, and Wiegand, Clyde E. "The Antiproton." *Scientific American* 194: 37; June 1956.
16 Sanford, Vera. *A Short History of Mathematics*. Boston: Houghton Mifflin Company, 1930. p. 3–4.
17. Piaget, Jean. *The Language and Thought of the Child*. 2nd edition. London: Routledge and Kegan Paul, Ltd., 1948.
18. Piaget, Jean. *Judgment and Reasoning in the Child*. London: Kegan Paul, Trench, Truebner & Company, 1928.
19. Piaget, Jean. *Judgment and Reasoning in the Child*. London: Kegan Paul, Trench, Truebner & Company, 1928. p. 138 ff.
20. Piaget, Jean. *Judgment and Reasoning in the Child*. London: Kegan Paul, Trench, Truebner & Co., 1928. p. 138.
21. National Council of Teachers of Mathematics. *Insights Into Modern Mathematics*. Twenty-Third Yearbook. Washington, D. C.: The National Council of Teachers of Mathematics, 1957. p. 146.
22. Heineman, E. Richard. *Plane Trigonometry*. 1st edition. New York: McGraw-Hill Book Company, 1942. p. 62.
23. Hendrix, Gertrude. "Developing a Logical Concept in Elementary Mathematics." *The Mathematics Teacher* 48: 542–47; December 1955.
24. Freeman, Archie, and others. "Symbolic Logic and Logical Circuitry in the High School." *The Mathematics Teacher* 50: 23–26; January 1957.
25. Allendoerfer, C. B., and Oakley, C. O. *Principles of Mathematics*. New York: McGraw-Hill Book Company, 1955.
Mathematics Staff of the College. *Concepts and Structure of Mathematics*. Chicago: The University of Chicago Press, 1956.
May, Kenneth O. *Fundamental Mathematics*. Preliminary edition 1955. Cambridge, Mass.: Addison-Wesley Publishing Company, 1955.
Stabler, E. R. *An Introduction to Mathematical Thought*. Cambridge, Mass.: Addison-Wesley Publishing Company, 1953

26. Courant, Richard, and Robbins, Herbert. *What is Mathematics?* New York: Oxford University Press, 1941. p. 11.
27. Allendoerfer, Carl B. "Deductive Methods in Mathematics." *Insights Into Modern Mathematics.* Twenty-Third Yearbook of the National Council of Teachers of Mathematics. Washington D. C.: The National Council of Teachers of Mathematics, 1957. p. 65.
28. Fawcett, Harold P. *The Nature of Proof.* National Council of Teachers of Mathematics, Thirteenth Yearbook. New York: Bureau of Publications, Teachers College, Columbia University, 1938.
29. Henderson, Kenneth B. "Hypotheses as a Point of Departure in Geometry Teaching." *The Mathematics Teacher* 40: 176–77; April 1947.
30. Henderson, Kenneth B. "An Experiment in Teaching Solid Geometry to Provide Training in Thinking." Unpublished Ph.D. Thesis. Columbus: Ohio State University, 1946.

5

Measurement and Approximation

JOSEPH N. PAYNE AND ROBERT C. SEBER

MANY MEASURES and numbers ... from the cradle (Weight: 6 lbs. 10 ozs.) to the grave (Age: 72 yrs., 6 mos., 3 days).

As a child grows he must learn to measure many aspects of his environment. He learns to associate a number with a given quantity. He must learn also to deal effectively with the numbers that he obtains from the measurement process. These are vital and necessary aspects of the education of every child.

This chapter contains three sections: (1) Estimation; (2) Measurement; (3) Approximation. Although the three topics are interrelated, each has aspects peculiarly its own. The section on estimation contains suggestions for teaching estimates of computations, of the number of objects, and of measurements. It is devoted mainly to specific suggestions for teaching estimation to elementary and junior high school pupils. Many of the ideas presented provide the background for the last section on approximation. Increasingly precise estimates provide a natural lead into a discussion of mathematical approximation. Estimates of measurements provide the lead into the second section, Measurement. The section on measurement deals with the nature of the measuring process and related ideas. Procedures are developed for computing with numbers arising from approximation. Section 3 deals with the way approximation can be treated mathematically. It should be of value to junior and senior high school teachers.

ESTIMATION

Which will Johnny choose? (Fig. 1) Johnny is a three year old confronted with the choice of one of the sets. If the objects are pieces of candy, he will likely choose the first; if they are bitter pills, he will likely choose the second. It may be that Johnny does not know the number of objects in either set and yet he knows that the first set contains more than the second, and that the second contains fewer than the first. By choosing both, he shows that the two sets combined are more than either; by choosing neither one, he shows that a set of none is less

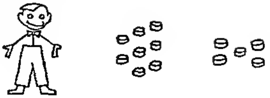


FIG. 1

than either of them. Early in their lives, children demonstrate some understanding of the idea of *more than* and *less than*. The idea is important mathematically and needs to be nurtured and cultivated throughout the years in school.

Often in arithmetic and in more advanced mathematics we are concerned with exact computation but neglect an equally important aspect, estimation. We may develop the algorisms for multiplication with understanding and yet fail to challenge pupils to think about the other relationships between numbers. We teach that 5×7 is 35 but leave the child to his own devices in realizing that 5 times a number more than 7 is a product more than 35. Though we teach him how to find the area of a rectangle and introduce the formula $A = lw$, we often overlook the highly useful aspect of having pupils estimate the length and width of a room to get a sensible estimate of the area of the rectangular floor. Without a design for learning which includes planned experiences in estimating and making sensible approximations, it is highly unlikely that students will achieve this by themselves. Certainly, estimating results of computations should become an integral part of the instructional program in arithmetic.

Mental and Paper-Pencil Estimation in Computation. A child's early idea of *more than* and *less than* is associated with physical quantities. Through visual observation, Tommy knows that he prefers the larger piece of cake, that he has more marbles than Jimmy, that he's smaller than his father, with little or no idea of how to number or measure objects in his physical environment. As he learns to count by ones to find the number of marbles or the number of pieces of candy, he focuses more attention on the number concept as an abstraction from a given set of objects. He learns the meaning of 5 from seeing and counting many sets of objects, each of which has one common characteristic, fiveness. As he makes comparisons he recognizes that a group of 8 contains more than a group of 5. Then, as his ideas expand, he begins to realize that each number is one more than the preceding one and that it is one less than the one following it. Furthermore, he learns that any number is more than *any* of those which come before it, e.g., that 8 is

greater than 7, 6, 5, 4, 3, 2, or 1. He discovers also that 8 is less than any number following it, i.e., 9, 10, 11, 12, $\cdots$. This early idea of *more than-ness* and *less than-ness*, the order relations, provides the early background for all further study of estimates of computations. It is used when students learn how to tell the larger of two fractions and then to make estimates of computations with fractions.

Early Elementary. In kindergarten and the lower elementary grades, teachers should provide frequent opportunity for pupils to estimate results of computations. After the pupils know that $3 + 4 = 7$, the teacher might ask if $3 + 6$ is more than 7 or less than 7. This question is based on the child's knowledge that 6 is more than 4. Later on as pupils learn the tens, the teacher could ask if $30 + 60$ is more than or less than 70. As multiplication is introduced, a similar procedure is appropriate. For example, knowing that $4 \times 3 = 12$, is 7×3 more than or less than 12? Thus, before the pupil knows the table of threes completely, he has some notion of the size of the product of 7 and 3. In a later grade, the teacher might ask if 3×70 is more than or less than 120.

The ability to estimate is extended when pupils are asked to give two numbers such that the result is between them. Knowing that $3 \times 4 = 12$ and $3 \times 10 = 30$, students can be led to see that 3×7 is between 12 and 30. The same idea is expressed when a pupil says that 3×7 is more than 12 but less than 30. In a division example like $31 \div 3$, pupils first might estimate the result as a little more than 10, basing their estimates on the known fact that $30 \div 3 = 10$. To bring in the idea of *between-ness*, pupils can locate two numbers such that $31 \div 3$ is between them. One pupil may know that $30 \div 3 = 10$ and that $36 \div 3 = 12$; hence, he may say that $31 \div 3$ is between 10 and 12, more than 10 but less than 12. Another pupil may think: $30 \div 3 = 10$; $33 \div 3 = 11$; hence $31 \div 3$ is between 10 and 11 but nearer 10 than 11. A similar analysis is appropriate for $32 \div 3$, $29 \div 3$, and $28 \div 3$. In locating two numbers such that the result is between them, we should expect some variation in pupil response, depending to some extent on the pupil's facility in making estimates.

Later Elementary. As a pupil progresses in arithmetic, place value as a characteristic of our number system becomes increasingly important. The position or place of the digit determines the value it represents; for example, in 368, the 3 means 3 hundreds, the 6 means 6 tens, and the 8 means 8 ones. This idea of place value becomes increasingly important in estimating computations also. Illustrations of the use of place value in estimating are shown when pupils are led to discoveries such as these: $310 \div 3$ is more than 100; 6×37 is between 180 and 240; 215×404

is about 80,000; 79,000,000 ÷ 2 is about 40,000,000 (pupils may see it more easily if written 79 million ÷ 2).

Estimating results can and should be extended to all parts of the instructional program in arithmetic. As complex fractions are studied, pupils can be asked to estimate sums such as $5\frac{3}{16} + 4\frac{5}{12}$. All pupils should be led to see that the result is between 9 and 11. Some pupils will be able to give a smaller interval such as *between 9 and 10*. Brighter students may see the result as a little more than $9\frac{1}{2}$ and may think like this: $5\frac{3}{16} + 4\frac{5}{16} = 9\frac{8}{16}$ or $9\frac{1}{2}$; since $4\frac{5}{12}$ is more than $4\frac{5}{16}$, $5\frac{3}{16} + 4\frac{5}{12}$ is a little more than $9\frac{1}{2}$. While estimating a sum like 12 + 14 + 9 + 18, some pupils will think: Each number is less than 20; hence the sum is less than 4 × 20 or 80. More observant youngsters will say that the sum is about 4 × 15 or 60.

Estimating products and quotients of decimal fractions helps pupils place the decimal point sensibly. For a product such as 3.24 × 4.6, students first should estimate; estimates should be in the interval 12 to 20 because 3 × 4 = 12 and 4 × 5 = 20. For example, students could be asked, "Estimate the product and then place the decimal point in the following."

$$\begin{array}{r} 3.24 \\ \times\ 4.6 \\ \hline 1944 \\ 1296 \\ \hline 14904 \end{array}$$

A similar procedure is appropriate for placing the decimal point in quotients. In an example such as 61.72 ÷ 4.2 students can make a preliminary estimate of 15 for the quotient. This should help pupils place the decimal point sensibly instead of following the rule blindly.

To make estimates of results involving decimals, pupils must recognize the relative size of numbers such as 3.24, 4.6, 61.72, and 4.2. Often pupils fail to do this when they erroneously conclude that the size of a number is determined by the number of digits. To help pupils gain clearer insight into the relative size of such numbers, it is helpful to have pupils count by tenths, hundredths, and thousandths. For example, to get a notion of the position of 3.2 in relation to 3 and 4, count by tenths from 3 to 4:

3.0, 3.1, *3.2*, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0.

The pupil can see that 3.2 is more nearly 3 than 4. A similar exercise is appropriate for comparing numbers like 62.3 and 62.27 by counting by hundredths from 62.20 to 62.30.

Pupils can learn to estimate results to obtain a reasonable check of their paper-pencil computations. Also, estimating leads directly to later work with slide rule calculations where one almost always relies on an estimate to place the decimal point in the result.

Students usually find it easy to count by 25's, 50's, and 100's. A practical use of such counting in making estimates is shown in an example from the grocery store.

Martha George is buying groceries in a super market. She comes to the candy counter and wonders if the \$5.00 she has is enough to pay for her groceries and buy a \$1.50 box of chocolates. To find out, she estimates the cost of groceries in her basket by rounding the cost of each item to the nearest 25 cents and and then mentally counts by 25's, 50's, or 100's, whichever is appropriate, to estimate the total.

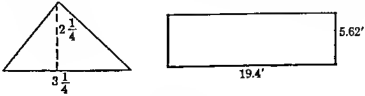
Actual cost		Cost rounded to nearest 25 cents	Mental Estimate (Cumulative total)
\$1.39		\$1.50	\$1.50
.37		.25	1.75
.15		.25	2.00
.98		1.00	3.00
.68		.75	3.75
\$3.57	Actual total		\$3.75

From this estimate of \$3.75, Martha George might decide that she should buy candy costing only \$1.00. With the totals cut from grocery slips, pupils can find the exact cost and compare it with their estimate. For variation, pupils can round each item to the nearest 10¢ and estimate the total by counting by tens.

Junior and Senior High School. As the formula for the circumference of a circle is introduced, pupils may bring in tin cans, bottle tops, and other circular objects, measure the circumference and the diameter and then discover that the circumference is *about* 3 times the diameter. Then the formula $C = \pi d$ is introduced. With the valid goal in mind of getting students to compute accurately and to use the formula properly, frequently we spend all the time on these computations, using $3\frac{1}{7}$ or 3.14 as an approximation of the number, π. It is equally important to help pupils extend their ability to estimate the results before and after the direct computation. Teachers should encourage pupils to make estimates while studying informal geometry, square roots, verbal problems, and applications in algebra and geometry.

While finding areas of triangles, rectangles, and other polygons, part of the time can be devoted to estimating the areas by estimating results

of computations. For example, students can be asked to estimate the area of the polygons such as those in Figure 2. For the triangle, an esti-



Estimate of area
FIG. 2

mate of 3 or 4 square units would seem reasonable. The area of the rectangle is about 100; some of the more able pupils might see the area as about 110 square units, obtained by multiplying 20 by 5½.

As pupils undertake square root in the junior high school, they should be encouraged to estimate before trying to find a closer approximation. Clearly, pupils should be led to see that $\sqrt{75}$ is between 8 and 9, that $\sqrt{175}$ is between 13 and 14, before proceding to find a closer approximation by the division method or the algorism. The division method for finding the square root is more meaningful than the usual algorism and should be given considerable emphasis, particularly as square root is introduced. This is true because the division method focuses attention on the fundamental notion of the square root of a number. The division method can be more than just trial and error. For example, a preliminary estimate of $\sqrt{75}$ might be 8.5.

Divide 75 by 8.5

```
        8.8
8.5/75.00
    68 0
     7 00
     6 80
       20
```

To make the second estimate, average 8.5 and 8.8, getting 8.65.

Use 8.65 as the trial divisor

```
          8.67
8.65/75.00
     69 20
      5 800
      5 190
        6100
        6055
          45
```

This process can continue until the desired precision is obtained. From the second trial division, it is clear that $\sqrt{75}$ is 8.7, to the nearest tenth and there is reasonable assurance that the result is 8.66, to the nearest hundredth. As the averaging and division process continues, a closer approximation is obtained although no decimal number will be the exact square root of 75. Teachers might point out that $\sqrt{75}$ is an irrational number. (See page 52 in Chapter 2, "Number and Operation.")

While working with formulas which require extensive computation, it is helpful first to estimate the result. To find the result for

$$\frac{31.72 \times 42.4}{167}$$

a pupil might estimate this way: "The result is about $(30 \times 40)/150$ or $(3 \times 40)/15$ or about $120/15$ or about 8." This kind of estimation may help avoid errors in placing the decimal point. Such exercises are also important in computations using the slide rule.

Often pupils use the rules blindly for finding the characteristic of the logarithm of a number. It is important that they be given ample practice in estimating the logarithm before extensive use of tables. For example, the log of 789 is between 2 and 3 because $10^2 = 100$ and $10^3 = 1000$.

As the reader has probably recognized, the more advanced work on estimates in calculations is dependent upon the ability to multiply and divide easily by 10, 100, 1000, and other powers of 10. Therefore, there should be continuing stress on multiplication and division by powers of 10 as well as stress on the structure of the numeration system.

Summary of Recurring Ideas. From the study of estimation in computation, certain ideas recur in varying degrees of importance from kindergarten through high school.

1. The order relations *greater than* and *less than* are important for almost all work on estimates of computations. The idea begins in kindergarten and lower elementary grades as pupils learn that every integer is greater than all that come before it and is less than all that follow it in counting order. In the middle and upper elementary grades this greater than-less than idea is extended to common and decimal fractions.

2. An idea of *between-ness* begins somewhat later than the early concepts of *more than* and *less than*. It represents a growth in the ability to estimate, i.e., the ability to establish two numbers such that the result is *between* them.

3. An understanding of the characteristics of our number system plays an important role in the growth of the ability to estimate. Place value is particularly important.

4. To estimate effectively, it is necessary to be able to multiply and divide easily by 10, 100, 1000, and other powers of 10.

5. Implicit in all work on estimation is the attempt to get pupils to think about the result and make an estimate before plunging directly into the exact paper-pencil computation. The major attempt is to get pupils to deal with numbers sensibly.

Estimating the Number of Objects. About how many people were at the meeting? About how many marbles are in the jar? What is your estimate of the number of children on the playground?

These questions suggest a second aspect of estimation, looking at a group of objects and estimating its number. The subject is often overlooked and many people are unable to make a reasonable guess in such situations. This section contains some suggestions for helping pupils grow in their ability to make estimates of this nature.

Early ideas in estimating the number of objects involve only more than, less than, many more, few more, and the like. Teachers can ask questions such as: Are there more people in that group than there are in our class? Many more? Few more? Then, as pupils grow in their ability to make such estimations, closer approximations may be expected. In the kindergarten and lower elementary grades, pupils are led to get a mental concept of groups of 1, 2, 3, 4, 5. Larger groups are usually visualized in terms of these smaller groups. For example, ten may be perceived as $5 + 5$ from the pattern

This same idea is used when we encourage pupils to estimate a larger group by estimating the number of smaller, more familiar groups it contains. For example, a pupil can count the number of people in a reading group, look at it to get some idea of the size, and then apply this *unit of measurement* to the total class and make some guess about the number of children. He can also count the number of children in the room and use this larger unit of measurement to estimate the number of children on the playground. To estimate even larger groups of people, students might be asked to focus attention on the number of people at assembly to get a larger unit for use in estimating the crowd in the football stadium.

Such estimates may involve mental or paper-pencil computations. To estimate the number of marbles in a jar, the child can count the number of marbles on the bottom of the jar, estimate the number of layers (by counting down the jar) and then multiply to get an estimate of the total

number. This could turn into an interesting guessing game. To help pupils estimate the number of seats in an auditorium or church, have them count the number of seats in one row, the number of rows, and then multiply mentally (or paper-pencil if too complicated) to find the total. Numerous opportunities arise where teachers can take a few minutes for this kind of estimation. At Thanksgiving and Christmas, with the abundance of cranberries, a guessing game is fun. The teacher might hold up a handful of cranberries and ask pupils to estimate the number.

The inverse of estimating a given number of objects is encountered when a number is given and a group of objects is to be chosen. "I want 100 sheets of paper." "Bring me a dozen pins." "I need 18 pieces of candy." Procedures for teaching this type of estimation are similar to those described earlier.

Estimating Measurements. The practical value of being able to estimate the size of physical objects is recognized by almost all people. "What is the length of this room?" "How far is it from the center of town to your house?" "How many feet were you from the scene of the accident?" "Does this letter weigh too much to go for 4 cents?" "How long should I allow for my drive to the airport?" "How much grass seed should I buy for my lawn?" "How many square yards of carpeting do I need?" A measuring instrument is not always handy or it may be difficult to measure the quantity directly. The teacher should realize that this kind of estimation is difficult and somewhat cumbersome to teach. There are several reasons for this. It is time consuming; it may be difficult to apply a measuring instrument to the quantity; and a certain degree of mathematical maturity seems desirable, if not necessary. But this does not mean that we should ignore this useful aspect of estimation. It is even more reason for giving concentrated and systematic study to this topic.

As with almost all measuring, we apply a known unit to an unknown quantity and determine the number of units. Hence, the essential idea of estimating measurements involves first-hand experience with a variety of units of measurement and relating these units to something familiar to the individual making the estimate. The need for giving pupils repeated experience cannot be overemphasized. A clerk at the meat counter, upon a request for a three pound roast, can take a large forequarter of beef and slice a roast within an ounce or two of the desired three pounds. He has had much experience in estimating.

Linear Estimates. To make linear estimates, each pupil can measure the breadth of his hand. Using the hand as a unit of measurement, he can be asked to find the width of several objects; desks, books, windows,

etc. Knowing the number of hands and the width of each hand, he can make an estimate. Similarly, pupils can get a smaller measuring unit by measuring a joint of a finger, perhaps trying to find one that is one inch long. Measuring his height and using this as a unit, a student might estimate that the height of the room is three times his height, or about 12 feet.

It is common to see the length of the room or the width of a lot *stepped off*. This is done by knowing the length of the stride, finding the number of steps, and then estimating the distance. Thus, pupils can measure their stride (perhaps take an average of 5 or 6 steps) and then practice using it. If desired, the distance may be measured with a tape measure or yardstick to check the estimate.

Estimating longer distances is not as simple because we cannot apply a known quantity as easily and must rely on a mental approximation of the number of units. It may be helpful to have pupils take the distance that they travel often and know well, e.g., distance from home to school, and try to apply this unit to other distances. "Well, it's about twice as far from my house to yours as it is from mine to school. It's 1.6 miles from my house to school. Hence, it must be about 3 miles from my house to yours." As a child begins to travel, larger units make more sense. It is 35 miles to City A. It must be about twice as far to City B; hence, it is about 70 miles.

The *Boy Scout Manual* contains many suggestions for making linear estimations, directly and indirectly. For further information on this topic, consult this reference.

Estimates of Other Measurements. Making estimates of other measurements involves procedures similar to those suggested for linear estimates. Area is usually estimated indirectly after measuring or stepping off linear measurements. The dimensions of a room may be estimated to be 21′ by 45′; hence, the area would be estimated as about 900 square feet. Likewise, most estimates of volume are made indirectly using known or estimated linear measurements. To help pupils estimate weight, they might be asked to estimate the weight of certain objects before and after handling an ounce and a pound weight. It is difficult to teach pupils to estimate the length of time from one event to the next. But they should be able to estimate driving time between two cities if the distance and average speed are known. Estimates of size of angles are usually made using the easily recognized 45°, 90°, 180°, and 270° angles as guides.

One practical problem that arises in estimations is the amount of error that should be allowed. To some extent the allowable error depends on

the use of the estimate. Certainly if a pupil can estimate with a 10 to 20% error, he is doing well. In the early work on estimation, we may expect much greater error. The study of the error of an estimate provides a nice opportunity for some work on per cent by finding the per cent of error of a given estimate. Likewise, pupils can calculate the amount of permissible error when the per cent allowed is stated.

We cannot expect a pupil to make satisfactory estimates of measurements from a brief exposure to the topic. Elementary and high school teachers should provide many and varied opportunities for pupils to make estimates of measurements. These experiences should be of the kind which relates estimation of measurements to the world and the environment of the student.

MEASUREMENT

The concept of measurement is used daily by everyone. It is so much a part of our everyday life that we are seldom aware of its extensive use. We get up at 7 a.m., allow 30 minutes for breakfast, drive our new 1959 car at a speed under the 55 m.p.h. limit. We encounter difficulties parking—our car is too long; there are too many cars and too few parking spaces. On a trip to the super market we see that each item is labeled with one or more number symbols referring to price, length, area, volume, and weight. We measure for draperies, rugs, and to find the amount of grass seed for our lawns. Meters measure the number of units of gas, water, and electricity we use. There is little doubt that measurement is an integral part of our lives.

This section is devoted to the teaching of direct and indirect measurement and related ideas such as standard units, precision, accuracy, tolerance, and tolerance interval. Considerable attention is given to procedures for making computations with numbers arising from approximations, including numbers arising from measurement.

Direct vs. Indirect Measurement. In every instance, the ultimate result of the measuring process associates a number with some geometrical or physical quantity. This association may be made directly or indirectly.

Direct measurements are made by a direct comparison of a unit of measurement with the object to be measured. Examples of direct measurement are: length of a room found by using a yardstick or tape measure; area of floor found by counting the number of tiles (square units) on the floor; volume within a bottle by counting the number of standard cups of water it takes to fill it.

When the measure of a quantity such as time, temperature, velocity,

weight, or even distance is computed from numbers obtained by actually measuring some other quantity, the first quantity is measured indirectly. Instruments such as clocks, thermometers, speedometers, and spring scales are used to make measurements indirectly. Such devices convert the quantity being measured into forces which move gears, expand metals, and in some way produce numbers which are read directly on a scale. Such measurements are made possible by combined mathematical, scientific, and mechanical knowledge. Some quantities are measured indirectly by using mathematical formulas to translate known direct measurements into a number for the quantity. The area of a rectangular floor is found indirectly when the number of units in the length and the number of units in the width are substituted in the formula $A = lw$. (If $l = 21$ ft. and $w = 10$ ft., then the area is found indirectly; $A = 21 \times 10$ or 210 sq. ft.) The height of the flagpole (Fig. 3) is found indirectly

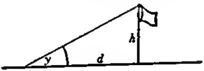


FIG. 3

by using the formula $h = d \cdot$ (tangent $\angle y$) and the direct measurements of angle y and d. (If angle $y = 41°$ and $d = 76$ ft., then $h = 76$ (tangent 41°); $h = 76$ (.869) or approximately $h = 66$ ft. In some cases we are unable to apply a measuring instrument directly to the quantity, e.g., finding the time which has passed since breakfast; in other cases, it is a matter of convenience to make measurements indirectly, e.g., the area of floor and the height of the flagpole.

Units of Measurement. From the study of measurement, pupils should realize that standard units of measurement are arbitrarily chosen although the units may have arisen historically. Furthermore, they should recognize that units of measurement are standardized to enable them to have widespread use. For example, a door made in Indiana must fit an automobile assembled in Michigan. To help achieve these objectives, a lesson such as the following might be planned. Have pupils place their elbows on their desks with the forearm pointed upward and the fingers extended. Measure the length from the elbow to the finger tip (this is the ancient *cubit*, a unit of length). Record the measurement for each pupil. They will recognize that a cloth merchant would profit by having a short arm. Recognizing the need for a standard unit, the class can establish its own by averaging the various measurements. Then, if each pupil is furnished a stick of this length, many measurements can be done with the unit standardized just for the class. If desired, the

class cubit can be divided into smaller equal units to measure smaller lengths.

Standard units of measurement are defined by law. The definitions of some units are made by the U. S. Congress and some are made by individual states. A bushel of corn may have different definitions in neighboring states. In most of our trading and industrial establishments we still use the English system of weights and measures established by act of Congress although the metric system is permissible. But the English units of measurement are defined by law in terms of the metric units which are generally used in scientific work and which are being used increasingly elsewhere. The yard is defined to be $\frac{3600}{3937}$ meter. This makes an inch a little more than 2.540005 centimeters. For industrial purposes the inch is often treated as exactly 2.54 centimeters. The definition of the meter itself, however, is in the process of being made more precise.*

Students should be familiar with the variety of units of measurement. Examples of some units can be exhibited and used in the classroom to give pupils repeated contact with the units and their use. Certainly every pupil should be thoroughly familiar with commonly used units such as inch, foot, yard, mile, ounce, pound, ton, and so on. But in the present atom and space age, there is a need for more attention to very small and very large units of measurement. Contrast in linear units can

* The following is quoted from *Science* 127: 76; January 10, 1958.

New Standard of Length

The Advisory Committee for the Definition of the Metre, chaired by L. E. Howlett, director of the Division of Applied Physics, National Research Council of Canada, has unanimously agreed on a new standard of length—a wavelength of light—to be used instead of the platinum-iridium bar kept at Sevres, France. The leading contenders as the source for light for the standard have been the following isotopes: mercury-198, krypton-84, krypton-86, and cadmium-114. One of the wavelengths of orange light emitted by krypton-86 has been selected as the standard, and the international meter will be defined as 1,650,763.73 times this wavelength. The resulting standard will be more than 100 times as precise as the present international meter.

Although in practice the new standard is already in use, several steps remain before the wavelength becomes officially recognized. The committee mentioned above will send its recommendation to the International Committee of Weights and Measures for consideration at its meeting in October 1958; when approved there, the recommendation will be presented to the International Conference on Weights and Measures, which will meet in 1960; at that time the standard will become the legal international standard.

be shown by tbese comparisons:

.00000001 cm.	—Diameter of hydrogen atom
.000001 in.	—Crosshairs for telescope made from cocoon of brown spider
.003 in.	—Thickness of human bair
.01 in.	—Thickness of 3 sheets of standard newspaper
5 light years	—Distance to one of the nearer stars (distance given in terms of time it takes a ray of light to reach earth).

Contrast in velocity can be shown by comparing the speed of a turtle to the speed of sound, about 1080 ft. per sec., or about $\frac{1}{5}$ mile per sec., the velocity of a satellite in orbit at about 18,000 miles per hour, and the speed of ligbt at about 186,000 miles per second, or about 670,000,000 miles per hour.

Students should be familiar with the many measuring instruments and realize man's quest for more precise instruments for measurement. Foot rulers, yardsticks, meter sticks, steel tapes, tape measures, vernier, and micrometer calipers are some of the instruments for measuring length which can be demonstrated in the classroom. Other instruments can be used and demonstrated such as the protractor and transit for measuring angles, and platform balances and spring scales for measuring weight.

Relationships Between Units. Letting generalizations grow from several concrete examples is commonly accepted as sound practice in teaching arithmetic. While teaching the basic addition facts, students are given repeated examples such as combining 3 beads and 4 beads, 3 chairs and 4 chairs, 3 pencils and 4 pencils before reaching the generalization, $3 + 4 = 7$. A similar procedure is higbly desirable for teaching many of the relationships between units of measurement. Students can be given a rule marked in foot units only and one with inch units only. By arranging interesting situations in which the student uses both rules to make the same measurement, it seems likely that he will make more observations about the relation between the foot unit and the inch unit. For 1 foot unit he has 12 inch units; for 2 foot units he has 24 inch units; for 3 foot units he has 36 inch units. He pairs the number of foot units with the number of inch units and generalizes that there are 12 inch units for each foot unit, 12 inches = 1 foot.

At a more advanced level, we can help the student see how he might record the results of his observations in the form of a table, generalize, and then write a formula expressing the relation.

number of foot units	1	2	3
number of inch units	12	24	36

Examination of this table might lead him to conjecture that when the number of foot units is 4, the number of inch units if 48. He might verbalize what he sees in the table, ***12 times the number of foot units is equal to the number of inch units.*** For the number of foot units, F, and the number of inch units, I, we write the condition that 12 times the number of foot units is equal to the number of inch units as $12F = I$. (Students should be reminded that the formula does *not* state that 12 feet = 1 inch; rather, 12 × the number of feet = number of inches.) Similarly he can discover or interpret the formulas, $3Y = F$, $36Y = I$, for the number of yard marks, Y, and the number of foot marks, F, and the number of inch marks, I. The idea of the relatedness of the number of one unit of measure and another unit of measure can be brought to fruition in a formula, illustrative of the way (with considerable simplification) in which a mature scientist examines a specimen of his environment, records pairs of numbers, and writes a formula to describe the results of his observations.

The relationship between units of weight such as ounce and pound can be shown by use of a platform balance. For liquid measure, pupils can measure using the cup, pint, quart, and gallon and then generalize about the relationship between the units.

By observing physical examples of units of measurement the student comes to know many of the directly observable units of measurement and the relationship between the same units. But an individual's immediate sensory perception does not permit him to know relationships such as those existing between one inch and one millionth of an inch, one mile and one light year, or one second and one century. How, then, can students come to know about the relation between such units? In these cases, we must rely on numerical processes. One inch = 1,000,000 millionths inches because 1 = 1,000,000 millionths. One light year is defined as the distance light travels in 1 year; hence 1 light year is approximately 186,000 × 60 × 60 × 24 × 365 miles. One century is approximately 60 × 60 × 24 × 365 × 100 seconds. This provides an opportunity to distinguish between relationships which are purely numerical (one contains 1,000,000 millionths), those which also involve physical or empirical relationships (the length of one light year depends upon the speed of light as well as the number of seconds in a year), and those which are *purely* definitional in character (there are 12 inches in a foot or an inch is 2.54 centimeters). This latter concept—that basic

units are arbitrarily defined—is a fundamental notion which must eventually be brought out clearly and not be obscured by the concrete approach suggested above as a pedagogical device for giving children an insight into the nature of units and their interrelationships. A clear understanding of this was noted in Chapter 4, "Proof," as one way of pointing out the fundamental role and arbitrary nature of definitions in all mathematical systems.

As has been pointed out earlier, by the measuring process we associate a number with some geometrical or physical quantity. Thus, measurement would include counting the number of fingers, the number of people in the room, or the number of dollars and cents in the bank. But the more common usage of measurement does not include these examples of a discrete number of objects. In the more common usage of measurement, a measuring instrument and a unit of measurement are implied. Furthermore, it is usually assumed that the quantity to be measured cannot be numbered exactly.

A clear conception of the measurement process must include an understanding of its approximate nature. In the early elementary grades, a pupil might think that he can cut off two feet of rope in the same sense that he can pick up two pennies. He measures perhaps without realizing that the length of rope is approximately 2 feet. But he will measure lengths such as one which is about $2\frac{1}{2}$ ft. where he is certain to observe that, "It does not come out even." He will see that it contains neither exactly two feet nor exactly three feet. There is nothing inexact about the amount of rope; there is a definite, fixed amount. There is nothing inexact about the number 2 or the number 3. But there is something inexact about the description of the length of rope as 2 ft. No matter how small the unit of measurement, the exact length cannot be determined. Thus, numbers arising from measurement made with rulers and other instruments are approximate representations of the physical quantity.

Precision of Measurement. *The precision of a measurement is defined by the unit used to make it.* When a measurement is made to the nearest foot, the precision of measurement is 1 ft. Measuring to the nearest half inch gives greater precision. The smaller the unit, the greater is the precision of the measurement.

The error of measurement is the difference between the actual length and the recorded length. Error of measurement does *not* refer to a mistake in measuring. If line segment AB (Fig. 4) is measured to the nearest half inch, the length is recorded as $2\frac{1}{2}$ inches. The error of measurement is the length of YB, the difference between the length of AB and the re-

FIG. 4

corded measurement of 2½ inches. When AB is measured to the nearest inch, the length is recorded as 3 inches and the error of measurement is $(3 - \overline{AB})$ inches. If AB is more than 3 inches, the error of measurement is $(\overline{AB} - 3)$ inches. It should be clear that the error of measurement cannot be determined just as the exact length of AB cannot be determined. We can only state the length of the interval within which the error must be contained.

To help pupils understand the meaning of precision and error of measurement the teacher can plan a lesson similar to the following. Distribute a sheet of paper to each pupil containing several line segments of different lengths. Also, supply students with a scale marked in one-inch units only. Have the students measure each line segment with the inch-unit scale and record their results to the nearest inch in a table. The pupils can be told that the error of measurement is the difference between the actual length and the recorded length. Then give the pupils a scale marked in one-half inch units and ask them to record the measurements. Then, using a scale marked in ¼ inch units only, repeat the process. If the original lines have been carefully drawn by the teacher, the student will obtain different numbers such as 3, 2½, 2¾ for successive measurements of the same line with different rulers! From such a laboratory lesson, the students should realize that *the smaller the unit of measurement, the more precise is the measurement.*

By pertinent questions from the teacher, the above lesson can be extended to help pupils understand *greatest possible error,* the tolerance, as it relates to measurement. When line segment AB is measured to the nearest inch, what is the recorded length? (Fig. 5). AB must be as long

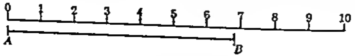


FIG. 5

as _?_ to be recorded as 7 inches. AB must be shorter than _?_ to be recorded as 7 inches Then, the greatest possible error is _?_. (Students will need a reminder that error is the difference between the actual length

and the recorded length.) By use of questions like these, pupils can see that the greatest possible error is ½ inch. Similar questions should be asked for measurement to the nearest half inch and nearest quarter inch. From such a discussion, pupils are led to make the added generalization: *The greatest possible error (the tolerance) is one-half the unit of measurement.*

Tolerance Interval. If any of the line segments AB CD, EF, or GH (Fig. 6) is measured to the nearest inch, the length will be recorded

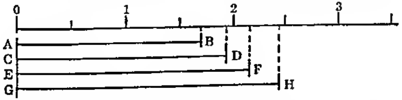


Fig. 6

as 2 inches. In other words, any length between 1½″ and 2½″ will be recorded as 2″ when measured to the nearest inch. The interval 1½″ to 2½″ is called the *tolerance interval* for a recorded measurement of 2 inches when made to the nearest inch. Thus when measuring with inch units, we do not know the number which represents the length of any one of the segments above. But we do know that the number for any one of them is in the tolerance interval 1.5 to 2.5. If another line segment XY is measured to the nearest half inch as 6½″, the number 6½ does not tell the exact length of XY; but we would know that the actual length of XY is in the interval 6¼″ to 6¾″. Thus, *in most cases, the tolerance interval is determined by the unit of measurement.* In such cases, the lower number of the tolerance interval is the recorded measurement minus one half the unit of measurement; the upper number is obtained by adding one half the unit of measurement. It is clear that in such cases, the tolerance interval is twice the tolerance.

In the previous examples, the units of measurement were stated and we were able to give the tolerance interval with certainty. In scientific and engineering work it is important and customary that these facts are written, read, and used carefully and properly. This is not always the case. Many times we must make an intelligent guess about the unit of measurement from a recorded measurement. For example, the mean distance to the sun is sometimes written as 93,000,000 miles; without knowing for sure about the unit of measurement, we might guess that it is *million miles* and infer that the tolerance interval is 92,500,000 miles to 93,500,000 miles. For a recorded length of 16.3 inches, we probably would assume that the unit of measurement is *tenth-inch* because

16.3 indicates this precision; if it were recorded as 16.30 inches, we would assume the measurement was made to the nearest *hundredth-inch*. Much of the work with measurement will require pupils to make a sensible guess about the unit from a recorded measurement. Hence, it seems reasonable to ask pupils to give the unit of measurement, the greatest possible error, and the tolerance interval for several measurements where the unit of measurement is not stated explicitly.

We should encourage pupils to inquire about the tolerance and the unit of measurement in each new measuring context. When a seamstress buys 3 yards of material, the clerk does not give her $2\frac{1}{2}$ to $3\frac{1}{2}$ yards of material. She is assured of getting at least 3 yards and the tolerance interval probably is 3 yards—3 yards 2 inches. If you go to a lumber yard and purchase a board 8 feet in length, what is the unit of measurement? What precision is required in milling this type of lumber? Does this mean that when you get home you can expect that the length of the board is more than 7.5 feet and less than 8.5 feet? If a man is working in a machine tool company and the blue print calls for a piece which is $\frac{3}{2}$ inches long, is he to assume that he will be performing his duty properly if he turns out a piece whose length is between $1\frac{1}{4}$ inches and $1\frac{3}{4}$ inches? No, he will either already have been informed about the required unit of measurement employed or will inquire about it for any new blue print on which it is not specified. The understanding that $\frac{3}{2}$ inches may mean 1.5 inches or 1.50 inches or 1.500 inches is important. In many cases the teacher can rely upon the local environment to provide industries from which the student can obtain information regarding the units of measurement used or the precision of measurement required. Students can be encouraged to inquire about the tolerance for a quart of milk, a gallon of gasoline, and so on, in their own state.

We should acquaint pupils with the standard ways in which precision of measurement is indicated. Most blueprints use the $\pm$ notation to indicate the tolerance and the unit of measurement. A measurement of 6 ft. $\pm\frac{1}{2}$ in. would mean that the tolerance interval is 5 ft. $11\frac{1}{2}$ in. to 6 ft. $\frac{1}{2}$ in. and that the unit of measurement is 1 inch. A bolt and a bolt hole both labeled as $\frac{1}{4}$ in. would not have the same tolerance; the bolt might have as its diameter $0.25 - .005$ in. making the tolerance interval 0.245 to 0.250, and the diameter of the bolt hole might be written as $0.25 + .005$ in. making its tolerance interval .250 to .255. Scientific notation is also used to indicate the precision of measurement. For example, 2.63×10^4 cm. would mean that the measurement was made to the nearest hundred centimeters. If it were recorded as 26,300 cm., there might be some doubt about the precision of the measurement. Likewise,

3.10×10^{-4} cm. would indicate precision to the nearest millionth-centimeter.

By examining many practical illustrations from science and industry, students can begin to appreciate the role which measurement and approximation play. In the mass production of many items—refrigerators, television sets, automobiles—which are produced on assembly lines, each part must be produced with its measure in a given tolerance interval in order that the parts will fit together and the final product will perform properly. From the planning stage, through production, and the checking of the quality of the product—an applied mathematical area itself—we are concerned about the tolerance interval and its relation to the final performance of the item.

COMPUTATIONS WITH NUMBERS ARISING FROM APPROXIMATIONS

Approximations arise from many sources. It has already been shown that numbers obtained by applying measuring instruments to a quantity are approximations. Approximations arise also as we attempt to write a decimal representation for numbers like $\frac{1}{7}$, π, $\sqrt{2}$, $\sqrt[3]{7}$, log 62, or sine 32°. Numbers such as 171,000,000, obtained by rounding some number to the nearest million, give rise to approximations. Computations with numbers arising from approximations should be done with the idea of their nature clearly in mind.

Many standard texts use *Approximate Numbers* to denote such numbers. Although this terminology is simpler to use, the writers of this chapter have chosen to avoid it because mathematically, numbers are not classed in this manner. Teachers who wish to continue the use of *Approximate Numbers* should not feel that they are committing a grave error. But they should be certain that students realize that this is a shortened and convenient terminology. No number is approximate. The '$5\frac{1}{2}$' in '$5\frac{1}{2}$ ft.' represents a number. This number, by itself, is never an approximation; hence it does not really make much sense to call it *exact* either. However, the number $5\frac{1}{2}$ may be used to represent the height in feet of a man, with the understanding that we do not know the number which represents his true height, but that $5\frac{1}{2}$ is close to it. The number $\sqrt{2}$ is not approximate but any decimal expression for $\sqrt{2}$ is an approximation to $\sqrt{2}$.

In the section which follows, procedures are developed for making computations with numbers arising from approximations. Most of the examples deal with numbers arising from measurement because of the

ease of illustration. The reader should be aware that the procedures should be used for approximations arising from any source.

Adding and Subtracting. What is wrong with each of these examples?

Example 1		Example 2	
Add	4.7 ft.	Subtract	6.497 in.
	3.824 ft.		3.5 in.
	8.524 ft.		2.997 in.

In Example 1, the measurement of 4.7 ft. indicates that it is made to the nearest tenth of a foot; thus, the actual length is as small as 4.65 or as large as 4.75. The measurement of 3.824 ft. indicates precision to the thousandths of a foot; thus, the actual length is as small as 3.8235 ft. or as large as 3.8245 ft. If the actual lengths were the smallest possible, the sum would be 8.4735 ft.; if they were the largest possible, the sum would be 8.5745 ft.

Smallest	Largest
4.65 ft.	4.75 ft.
3.8235 ft.	3.8245 ft.
8.4735 ft.	8.5745 ft.

The sum of the two measurements is somewhere in the interval 8.4735 to 8.5745. The chances are only 1 in 100 that 8.524 (the sum recorded in Example 1) is the actual sum of the two measurements, to the nearest thousandth of an inch.

In adding the two measurements in Example 1, the precision of one of them, 4.7 ft., has been implicitly increased by adding two zeros getting 4.700 ft., indicating precision to the thousandth-inch. The precision of a measurement cannot be increased in this manner, merely with a flick of the pencil. To make it more precise, a smaller unit of measurement must be used.

The folly of such additions can be shown by placing cross marks in the spaces:

Example 1		Example 2	
Add	4.7xx ft.	Subtract	6.497 in.
	3.824 ft.		3.5xx in.

Not knowing what digits occupy the spaces, it is evident that the sum in Example 1 and the difference in Example 2 cannot be obtained to the nearest thousandth-inch.

From the previous discussion the following general principle is evident: *To add or subtract numbers arising from approximations, first round each number to the unit of the least precise number and then perform the operation.*

The principle stated above is easy to apply when measurements are recorded in the decimal system or in the metric system. It is not as easy when adding measurements such as 5¾ pounds and 6⅝ pounds. The main difficulty arises in assuming that 5¾ pounds represents a weight to the nearest quarter-pound; in actual practice it may mean 5⅝ pounds, a measurement to the nearest eighth-pound. Thus, in actual practice, you need to know the unit of measurement before you can perform the computations with certainty of obtaining the correct precision in the result.

Relative Error and Accuracy. If the length of a desk is recorded as 3½ feet to the the nearest half-foot, the greatest possible error is ¼ foot and the relative error is the ratio of ¼ to 3½ or ¼ ÷ 3½ which is .071. The *relative error* of a measurement is defined as *the ratio of the tolerance or maximum error to the measured value.* Some persons in some situations choose to express this fraction as a per cent (7.1 per cent in the above example). This is then called the *per cent of error.* By definition a measurement with a smaller relative error is said to be more *accurate* than one with a larger relative error. This is a special technical use of the words *accurate* and *accuracy.* It does not refer to the care with which the measurement was made; it is assumed that all measurements are made with care and correctly to the precision indicated for them.

It is not particularly easy to get pupils to realize the full impact of the statements in the previous paragraph. To introduce the idea, the teacher may give several sets of measurements and ask pupils to find the error and relative error as in the following examples. (In a work sheet given to students, the error column would be blank as is the relative error column.)

		Error	Relative Error
Example 1	12 ft.	½ ft.	——
	18 ft.	½ ft.	——
	6 ft.	½ ft.	——

Example 2	50 ft.	5 ft.	——
	50 yds.	5 yds.	——
	510 miles	5 miles	——
Example 3	1,000,000 miles	500,000 miles	——
	1 mi.	½ mi.	——
	.5 ft.	.05 ft.	——

From examples like these, pupils can be led to see that the accuracy of a measurement is independent of the unit of measurement used to make it.

To teach an understanding of this have students examine the following calculations of the relative error of the measurements 640 ft. to nearest 10 ft., 64 ft., 6.4 ft., and .64 ft.

$$\text{Relative error of 640 ft.} = \frac{5}{640} \qquad \text{or about .07}$$

$$\text{Relative error of 64 ft.} = \frac{.5}{64} = \frac{5}{640} \qquad \text{or about .07}$$

$$\text{Relative error of 6.4 ft.} = \frac{.05}{6.4} = \frac{5}{640} \qquad \text{or about .07}$$

$$\text{Relative error of .64 ft.} = \frac{.005}{.64} = \frac{5}{640} \qquad \text{or about .07}$$

Note, the relative error of each measurement is the same, about .07.

By similar calculations, note that the relative error of each of the measurements 365 in., 36 5 in., 3.65 in., and .365 in. is the same, approximately .0014. Hence each of this last set of measurements is more accurate than any of the first set of measurements. Each of the measurements 640 ft., 64 ft., 6.4 ft., .64 ft. has 2 digits which affect the accuracy of the measurement; each of the measurements 365 in., etc., has 3 digits which affect the accuracy of the measurement. Digits which affect the accuracy of a measurement are called *significant digits*. The more significant digits a measurement has, the more accurate is the measurement.

From the discussion above, it should be clear that *all nonzero digits are significant*. Zeros are sometimes significant; they are significant when they are between nonzero digits as in numerals like 700.01, 60054, and 3.005. Zeros are not significant in a numeral such as 0.004 because the zeros do not affect the relative error and hence the accuracy of the

measurement. For example, let's compare the relative error of .004 with that of .4 under the assumption that in each case the tolerance is one half a unit of the size represented by the position of the fours. Thus the relative error of $.004 = \frac{.0005}{.004} = \frac{5}{40} = .125$ while the relative error of $.4 = \frac{0.05}{.4} = \frac{5}{40} = .125$. This shows that the zeros in .004 are not "significant."

In a numeral such as 64,000, it is difficult to determine the number of significant digits because the unit of measurement is not known. If it is 64,000 to the nearest thousand, then none of the zeros is significant. If it is measured to the nearest unit, all the zeros are significant. To avoid ambiguity in cases such as this one, a dot can be placed over the zero which indicates the precision of the measurement. For example, $97{,}000{,}\dot{0}00$ would indicate measurement to the nearest hundred and there would be 6 significant digits. A second and perhaps more common way of representing significant digits in this case is to write the number in *standard* (sometimes called *scientific*) notation. In this system 97,000,000 to the nearest hundred would be written 9.70000×10^7. It should be clearly recognized that the number of significant digits is only a rough index of the accuracy of a measurement. Significant digits are introduced to facilitate the procedures for the other operations (multiplication, division, and square root) with numbers arising from approximations.

Multiplying and Dividing. Morgan is a model airplane enthusiast and has several gasoline planes. He flies the planes by attaching a piece of string to a plane and, by holding one end, lets the plane fly in a circular path. Morgan wanted to calculate the distance the plane travelled in one lap. He measured the diameter of the circle to be 68 feet. He knew that $C = \pi D$, and that π is approximately 3.1416. He found the circumference of the circle like this:

$$\begin{array}{r}
3.1416\phantom{\text{ feet}} \\
68\text{ feet} \\
\hline
25\ 1328\phantom{\text{ feet}} \\
188\ 496\phantom{8\text{ feet}} \\
\hline
213.6288\text{ feet}
\end{array}$$

Is his calculation correct? Is the circumference of the circle 213.6288 ft.? After measuring the diameter to the nearest foot (68 ft.) do you think that he could calculate the circumference to the nearest ten-thousandth of a foot?

Dan reminded Morgan that his measurement of 68 feet is made to the nearest foot and hence the actual diameter of the circle is between $67\frac{1}{2}$ feet and $68\frac{1}{2}$ feet. Furthermore, he noted that a 3 digit approximation for π is more sensible to use. Then Dan showed Morgan that the circumference of the circle could be as little as 211.950 feet and as much as 215.090 feet. He showed Morgan these calculations:

With Diameter of 67.5 ft.	With Diameter of 68.5 ft.
67.5 feet	68.5 feet
3.14	3.14
2700	2740
675	685
2025	2055
211.950	215.090

Morgan could have been asked to compare the digits in corresponding places in the two products. The digits are the same, 2 and 1, in the hundreds and tens columns, but different in the other places. Hence, it is clear that the result should be rounded to 210 ft. because there is no certainty of any digit to the right of the tens place.

Perhaps you recognize the important principle that *the accuracy of a measurement cannot be improved by computation.* That is to say, Morgan couldn't get a measurement of the circumference of a circle to the nearest ten-thousandth of a foot by using a five digit number for π, if his measurement originally had only two digit accuracy. He couldn't improve his accuracy (.007 relative error in 68 ft.) by multiplying. He merely put on an appearance of an accuracy to which he was not entitled when he used 3.1416 for π and wrote his result as 213.6288 feet.

A rough guide for rounding an answer obtained by multiplying numbers arising from approximations is as follows (remember that the number of significant digits gives a rough idea of accuracy):

In multiplying numbers arising from approximations, keep as many significant digits in the product as there are in the number with the fewer significant digits.

When Morgan found the circumference of the circular path which the plane travelled, he multiplied 3.1416 by 68 feet. The number 68 feet has the fewer significant digits (2) and hence he should have rounded his answer to contain only 2 significant digits (210 ft.).

The folly of keeping more significant digits in the answer is illustrated by placing an x (cross mark) for the digits which are not known and then attempting the multiplication.

```
   3.14
   68.x
   ____
   xxxx
  2512
 1884
 ______
 21x.xxx
```

Since the digit in the tenths place of 68 ft. is not known, we can indicate it by a cross mark, carry out the calculation and get 21x ft. which says that the result should be written as 210′ with two significant figures.

A geometric illustration of the need for rounding the result may be shown using a rectangle. Consider a rectangle (Fig. 7) whose length is

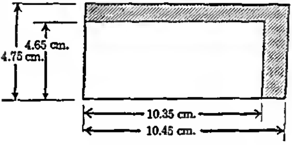


FIG. 7

recorded as 10.4 centimeters and width as 4.7 centimeters, with each measurement given to the nearest millimeter (tenth-centimeter).

Two of the possible rectangles are drawn with the measurements of each within the tolerance intervals. Of course, the actual rectangle may have two of its sides anywhere in the shaded area.

The area of the rectangle can be as little as 4.65×10.35 or 48.1275 square centimeters. The area can be as much as 4.75×10.45 or 49.6375 square centimeters. Thus, *the actual area is in the interval* 48.1275 *to* 49.6375 square centimeters.

Following the rule on multiplying numbers arising from approximation, we obtain the following result for the area of the rectangle, 10.4 cm. by 4.7 cm.

```
  10.4 cm.
   4.7 cm.
  ____
  7 28
 41 6
 ______
 48.88 or 49 sq. cm.
```

Dan reminded Morgan that his measurement of 68 feet is made to the nearest foot and hence the actual diameter of the circle is between 67½ feet and 68½ feet. Furthermore, he noted that a 3 digit approximation for π is more sensible to use. Then Dan showed Morgan that the circumference of the circle could be as little as 211.950 feet and as much as 215.090 feet. He showed Morgan these calculations:

With Diameter of 67.5 ft.	With Diameter of 68.5 ft.
67.5 feet	68.5 feet
3.14	3.14
2700	2740
675	685
2025	2055
211.950	215.090

Morgan could have been asked to compare the digits in corresponding places in the two products. The digits are the same, 2 and 1, in the hundreds and tens columns, but different in the other places. Hence, it is clear that the result should be rounded to 210 ft. because there is no certainty of any digit to the right of the tens place.

Perhaps you recognize the important principle that *the accuracy of a measurement cannot be improved by computation*. That is to say, Morgan couldn't get a measurement of the circumference of a circle to the nearest ten-thousandth of a foot by using a five digit number for π, if his measurement originally had only two digit accuracy. He couldn't improve his accuracy (.007 relative error in 68 ft.) by multiplying. He merely put on an appearance of an accuracy to which he was not entitled when be used 3.1416 for π and wrote his result as 213.6288 feet.

A rough guide for rounding an answer obtained by multiplying numbers arising from approximations is as follows (remember that the number of significant digits gives a rough idea of accuracy):

In multiplying numbers arising from approximations, keep as many significant digits in the product as there are in the number with the fewer significant digits.

When Morgan found the circumference of the circular path which the plane travelled, he multiplied 3.1416 by 68 feet. The number 68 feet has the fewer significant digits (2) and hence he should have rounded his answer to contain only 2 significant digits (210 ft.).

The folly of keeping more significant digits in the answer is illustrated by placing an x (cross mark) for the digits which are not known and then attempting the multiplication.

```
   3.14
   68.x
   ----
   xxxx
  2512
 1884
------
21x.xxx
```

Since the digit in the tenths place of 68 ft. is not known, we can indicate it by a cross mark, carry out the calculation and get 21x ft. which says that the result should be written as 210′ with two significant figures.

A geometric illustration of the need for rounding the result may be shown using a rectangle. Consider a rectangle (Fig. 7) whose length is

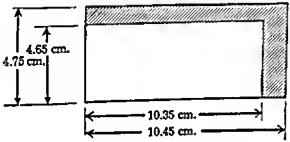


FIG. 7

recorded as 10.4 centimeters and width as 4.7 centimeters, with each measurement given to the nearest millimeter (tenth-centimeter).

Two of the possible rectangles are drawn with the measurements of each within the tolerance intervals. Of course, the actual rectangle may have two of its sides anywhere in the shaded area.

The area of the rectangle can be as little as 4.65 × 10.35 or 48.1275 square centimeters. The area can be as much as 4.75 × 10.45 or 49.6375 square centimeters. Thus, the actual area is in the interval 48.1275 to 49.6375 square centimeters.

Following the rule on multiplying numbers arising from approximation, we obtain the following result for the area of the rectangle, 10.4 cm. by 4.7 cm.

```
 10.4 cm.
  4.7 cm.
 -----
 7 28
41 6
-----
48.88 or 49 sq. cm.
```

Since 4.7 centimeters has the fewer significant digits (2), the result is rounded to 49 square centimeters. Writing our result this way implies that it is correct to the nearest square centimeter. The tolerance interval which we computed was 48.1275 to 49.6375 rather than 48.5 to 49.5. This shows that our rule about rounding products to the same number of significant figures as in the least accurate factor is itself only approximately correct. However, it is a good approximation and saves us from two long multiplications.

Division is the inverse of multiplication. Therefore, it would seem reasonable to follow the procedure used for multiplication in making divisions involving numbers arising from approximations.

In dividing numbers arising from approximations, keep as many significant digits in the quotient as there are in the number with the fewer significant digits.

A similar rule applies to square root. For example, if a square contains 69 square inches, measured to the nearest square inch, to find the side, take the square root of 69 and retain two significant digits in the result. Thus the side of a square whose area is 69 square inches is 8.3 inches, correct to two significant figures..

To find the square root of a number arising from approximation, retain as many significant digits in the result as the number contains.

APPROXIMATION

A physical conception of approximation and successive approximation arises with a measuring process and the use of smaller and smaller units of measurement. A formal notion of approximation grows from and brings new meaning to a student's work with estimations and measurements. In each of these we found as a key idea an ordering of quantity and of numbers. The idea of order continues in this role as we attempt to give it depth and breadth by viewing it in a variety of new contexts. One of these is approximation.

The continued study of mathematics demands that more and more attention be given to a distinction between mathematical conditions and the physical situations which they describe. Even though this distinction is made, there is little doubt about the pedagogical value of keeping both in view when dealing with either. In teaching geometry this point is illustrated when we help the student see a mathematical proof in terms of a sequence of deductions made from assumed conditions. A drawing may suggest that a statement is true. Other drawings may provide evidence that the statement is true, but they do not constitute a mathematical proof. A mathematical proof exhibits relations involving terms and statements. A student's understanding of these

terms and statements is extended by considering these relations. We help him see approximation in this way as conditions are developed which define an approximation to a number or a point. In doing this we find that algebraic and geometric conditions correspond to one another. Graphical and other physical interpretations of these conditions provide the opportunity to further develop an understanding of the distinction between the use of mathematical conditions in deductive schemes and the related physical concepts.

In this section you will find some of the contexts in which we make use of rational approximations. We find a concept of approximation involved as a student extends his notion of number from fraction or rational number to irrational number. We find approximations necessary when we make use of trigonometric or logarithmic tables and when we interpolate in them. Work with sequences of numbers brings into play a concept of successive approximation which provides part of the background for the study of limits and calculus. A geometric view of approximation is developed and related to a number line. This is indicative of the way in which we associate algebraic and geometric concepts. We find that approximation is not a property of a number but a relation involving pairs of numbers. The relation *is approximately* or "$\approx$" is defined consistently with the use of the term 'relation' in Chapter 3. Finally, these ideas are brought to bear on a brief discussion of some types of approximations which arise in science.

Rational Approximation. Children encounter a notion of approximation early in their school days when they first meet the division algorism. In dividing 10 by 4 the student can view the information obtained from the first step in the algorism in a variety of ways. One way to view it is that it gives the first, or in this case the units, digit in the quotient. But he can also state that this tells him that 2 is less than the quotient and the quotient is less than 3, or the quotient is between 2 and 3, or a first approximation to the quotient is 2, or the *quotient is approximately 2*. The recognition that these statements about the quotient are equivalent is important as we proceed in the direction of an algebraic or geometric treatment of approximation. Thus the division algorism can serve as a basis for discussions of decimal approximations to rational numbers.

One way in which we develop meaning in mathematics is by helping students recognize the common elements in different situations and different approaches to or symbols for the same thing. In our example where the division algorism is used to divide 10 by 4 we use the algorism to justify the equality of 10/4 and 2.5.

The use of pairs and triples of numbers and the relations which are

defined by sets of them is common to the study of many mathematical systems. Hence, emphasizing pairs and relations also helps teach meanings. An approximation relation arises naturally in connection with the use of the division algorism. It seems reasonable to use the algorism to determine a decimal representation of 10/3. In making use of the algorism we find that at each step the remainder is not zero. This is not like the case where the algorism is used to divide 10 by 4. The advantage of viewing the algorism in a variety of ways is more apparent here. We cannot conclude that 10/3 is equal to 3.3. However, the student relates 10/3 and 3.3 through the use of the algorism. As teachers we should encourage him to consider the nature of this relation by helping him draw conclusions about 10/3 and 3.3 from his work with the division algorism. He should see that the first two digits in a decimal representation of 10/3 are 3 and 3; 10/3 is between 3.3 and 3.4; 3.3 is less than 10/3 and 10/3 is less than 3.4; 10/3 is approximately 3.3.

These statements concerning the fraction represented by '10/3' can be written in symbolic form as below.

3.3 is less than 10/3 and 10/3 is less than 3.4

$$3.3 < 10/3 \quad \text{and} \quad 10/3 < 3.4$$

10/3 is between 3.3 and 3.4

$$3.3 < 10/3 < 3.4$$

10/3 is approximately 3.3

$$10/3 \approx 3.3$$

Many different problems involving division provide the opportunity for the teacher to help the student discover that there are numbers which cannot be represented in decimal notation by a finite sequence of digits, and that sets of numbers represented by finite sequences of digits are used to approximate these numbers such as 3.3, 3.33, 3.333, ··· and so on, which are all terminating decimals each of which is a closer approximation to 10/3 but none of which equals 10/3. In junior high school the concept of approximation which arises in connection with the use of the division algorism and in working with fractions which cannot be represented in decimal notation by a finite sequence of digits serves as a background from which the student can continue to explore mathematically the notion of approximation. To do this, however, he must begin to see the variety of mathematical statements which are equivalent to one another.

It is important for the student to associate the notion of approxima-

tion with a concept of order for numbers. The teacher can help him do this by (1) using a symbol for approximation when it is appropriate, (2) using the symbol which indicates that one number is less than another number, (3) bringing out in class discussion that there may be many different approximations to a given number, (4) relating the concepts of order and approximation by examples similar to the one above, and (5) developing the idea that approximation is a relation between numbers by exhibiting pairs of numbers in this relation. Each of these points will be illustrated again as we consider rational approximations to irrational numbers.

Rational Approximation to Irrational Numbers. The junior high school mathematics program serves to help the student make the transition from arithmetic to the algebra and geometry of the high school. Consequently we find many topics in the junior high school program in which the student can begin to make use of the idea of formulating and using mathematical conditions in his work. His work with informal geometry, formulas, equations, and graphs can lead to many interesting discoveries. Here we suggest ways in which a concept of approximation can be developed from a geometric situation. Informal geometry provides numerous situations in which we use numbers which cannot be represented by a finite sequence of digits. In the formulas for the area or circumference of a circle we find the number π. The length of the diagonal of a unit square is $\sqrt{2}$. These numbers are related by approximation to numbers represented by finite sequences of digits.

In addition to the name of a geometric configuration, such as a triangle, we consider properties of the configuration. For a right triangle we find that the area of the square constructed on the hypotenuse is equal to the sum of the areas of the squares constructed on the legs of the triangle. If A is a unit square (Fig. 8) and B is a unit square, we

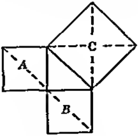


FIG. 8

can demonstrate that the area of square C is the sum of the areas of squares A and B. Figure 9 indicates one way in which this can be done This type of investigation provides some physical evidence that there

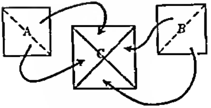


FIG. 9

is a square whose area is equal to that of two unit squares. Now one wonders what the length of a side of square C is. We might arrive at the same question by considering the formula $A = s^2$, for the area of a square. If 'A' is replaced by '2', is there a number for which $2 = s^2$ is true? One way to answer this question might be to construct a right triangle whose legs are each one unit long, then measure the length of the hypotenuse. This process may not prove to be very satisfactory. Suppose on the basis of measurement it is suggested that the length of a side of square C is 1.4 units. Since $(1.4)^2 = 1.96$ we find that 1.4 does not satisfy the condition $2 = s^2$. If the difficulty lies in the approximate nature of the measuring process and we cannot obtain an answer to our question by measuring, how can we answer it? This type of situation provides an excellent opportunity to illustrate how mathematical conditions can be used to obtain answers to questions.

What would be true of a number whose square is 2? The geometry may suggest the observation:

1 is less than the number and the number is less than 2,

which can be written also as

$$1 < n \quad \text{and} \quad n < 2.$$

In addition to the condition that the square of a number is equal to two, we have a condition which gives its order in the number system. For the condition '$1 < n$ and $n < 2$' we can consider replacements for 'n' for which the condition is true. Students may suggest some numbers, from the set of numbers with which they are familiar, for which the condition '$1 < n$ and $n < 2$' is true. Is one of these a number whose square is 2? Suppose that the candidates are 1.2, 1.41, 1.5032, 1.76, 1.52, 1.42, 1.4. In this manner we focus the student's attention on the set of replacements for 'n' for which the condition '$1 < n < 2$' is true. Checking to find out if any of these numbers has a square which is 2 we find that some of these numbers have squares less than 2 and the others have squares greater than 2.

This procedure is suggestive of the way in which we are lead to consider the condition '$1.41 < n$ and $n < 1.42$' in place of the condition

'$1 < n$ and $n < 2$'. The condition '$1.41 < n < 1.42$' is not true when 'n' is replaced by 1.2 or 1.76. The set of numbers for which '$1.41 < n < 1.42$' is true does not contain 1.2, 1.76, 1.52, 1.4, 1.5. However, 1.4109, 1.4137, 1.418, 1.4199 belong to this set.

We have dealt at length with this particular example because it is indicative of the manner in which we can begin to lay foundations for a formal development of the concept of approximation and successive approximation. We find that in addition to considering a number, we can encourage students to think about sets of numbers, such as the set of numbers between 1.41 and 1.42 or the set of numbers for which the condition '$1.41 < n < 1.42$' is true. We can use mathematical conditions in the discussion of approximation. Work of this type in the classroom can help the student begin to appreciate mathematics as more than a collection of problem solving techniques.

Even though the student is not familiar with irrational numbers, the geometric context in which the question of the existence of a number for which the condition '$s^2 = 2$' is true makes it seem reasonable that there is such a number. We must help the student understand that this number is not a rational fraction. It is not in the set of numbers with which he is familiar, but it can be approximated by rational numbers. These ideas can be developed by guiding attempts to produce a number whose square is 2 into the formulation of conditions similar to:

$$1 < n < 2,$$

$$1.4 < n < 1.5,$$

$$1.41 < n < 1.42,$$

$$1.414 < n < 1.415.$$

Some guessing, computing, and the recording of the results of our observations as mathematical conditions are fruitful types of class activity. They provide meaning for concepts before the introduction of symbols, technical terms, or formal definitions and rules. In the case of rational approximations we want the student to understand that $\sqrt{2}$ is approximately 1.41, $\sqrt{2} \approx 1.41$. This latter may be interpreted to mean that $\sqrt{2}$ is between 1.405 and 1.415, but in this case we have by earlier computation also established that in fact $\sqrt{2}$ is between 1.41 and 1.42. These statements can be written symbolically as

$$1.405 < \sqrt{2} < 1.415,$$

$$1.41 < \sqrt{2} < 1.42,$$

$$1.41 < \sqrt{2} < 1.415.$$

We also want to help the student develop a notion of successive approximation. The idea that $\sqrt{2} \approx 1.4$, $\sqrt{2} \approx 1.41$, $\sqrt{2} \approx 1.414$ suggests that there is a set of numbers such that each number in the set is related to $\sqrt{2}$ by approximation. The numbers which belong to this set will be determined by the conditions we use to define approximations to $\sqrt{2}$. Assume that we have developed the idea that $\sqrt{2}$ satisfies each of the conditions:

$$1 < n < 2$$

$$1.4 < n < 1.5$$

$$1.41 < n < 1.42$$

$$1.414 < n < 1.415$$

or that $\sqrt{2}$ is in each of the sets

$$\{n \mid 1 < n < 2\}$$

$$\{n \mid 1.4 < n < 1.5\}$$

$$\{n \mid 1.41 < n < 1.42\}$$

$$\{n \mid 1.414 < n < 1.415\}.$$

We might inquire about the numbers in the set $\{n \mid 1 < n < 2\}$. The first digit in the decimal representation of each number in this set is 1. $\sqrt{2}$ is in the set $\{n \mid 1 < n < 2\}$ since $\sqrt{2}$ satisfies the condition $1 < n < 2$. Therefore the first digit in the decimal representation of $\sqrt{2}$ is 1. We may treat each number in the set $\{n \mid 1 < n < 2\}$ as an approximation to $\sqrt{2}$. We can define an approximation to $\sqrt{2}$ as a number which satisfies the condition $1 < n < 2$. However it is important that we may also give other definitions. $\sqrt{2}$ is also in the set $\{n \mid 1.4 < n < 1.5\}$. The first two digits in the decimal representation of each number in this set are 1 and 4. Therefore the first two digits in the decimal representation of $\sqrt{2}$ are 1 and 4. We may define an approximation to $\sqrt{2}$ as a number which satisfies the condition $1.4 < n < 1.5$.

Now we may compare these two definitions:

(I) An approximation to $\sqrt{2}$ is a number which satisfies the condition

$$1 < n < 2.$$

(II) An approximation to $\sqrt{2}$ is a number which satisfies the condition

$$1.4 < n < 1.5.$$

Each number which is an approximation to $\sqrt{2}$ using definition (II) is also an approximation to $\sqrt{2}$ using definition (I). Not all the numbers which are approximations to $\sqrt{2}$ using definition (I) are approximations to $\sqrt{2}$ using definition (II). In this sense we say that the second definition gives closer approximations than the first definition.

In addition to representing a number whose square is 2 by $\sqrt{2}$ we gain insight into the nature of this number as we consider its position in the number system through decimal approximations to it.

$$\sqrt{2} \neq 1.4, \quad \text{but} \quad 1.4 < \sqrt{2} < 1.5$$
$$\sqrt{2} \neq 1.41, \quad \text{but} \quad 1.41 < \sqrt{2} < 1.42$$
$$\sqrt{2} \neq 1.414, \quad \text{but} \quad 1.414 < \sqrt{2} < 1.415$$

Beginning in junior high school and continuing through high school and college work in mathematics the student makes use of a variety of symbolic forms to represent numbers. He encounters 'π', '$\sqrt{2}$', 'sin 22°', 'log 37', each representing a number. A concept of approximation in terms of the first two, three, four, or five digits in the decimal representation of each of these numbers relates the number to a rational number.

$$\pi \approx 3.14, \qquad \sqrt{2} \approx 1.414, \qquad \sin 22° \approx .37461, \qquad \log 37 \approx 1.56820$$

Here we find the early notion of place value in the decimal representation of a number extended to thinking of numbers being represented by sequences of digits. Some of these sequences are finite, some are not. When a number is designated by an endless sequence of digits we relate it to a number represented by a finite sequence of digits by approximation. When a student uses tables such as a five place trigonometric table he should be helped to understand that the table gives the first four digits in the sequence of digits representing sin 22°. The fifth digit indicates the order of sin 22° in the number system. From the entry .37461 corresponding to sin 22° in a trigonometric table we know that the first four digits in the sequence representing sin 22° are 3, 7, 4, and 6. The fifth digit, 1, indicates that

$$.374605 < \sin 22° < .374615$$

and thus that the fifth digit may be 0 or 1.

Geometric Approximation. One of the fascinating aspects of mathematics is the variety of vantage points from which one can view a topic, each providing something in itself and still contributing to the others. One of the goals of instruction in high school mathematics is

to help the student begin to understand the role of definitions and deductive organization in mathematics. We should not overlook the opportunity to utilize the students' experiences with rational approximation as we move in the direction of formal definition and proof. This section contains a discussion of approximation in a geometric context which is suggestive of the way in which an order of the points belonging to a line is associated with an ordering of numbers.

We think of an ordering of the points belonging to a line corresponding to our notion of an ordering of numbers. We indicate this by saying that for points P and Q in Figure 10 belonging to a line, P precedes Q. The line segment with end points A and B, where A precedes B (Fig. 11)

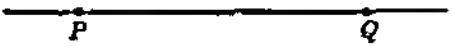


FIG. 10

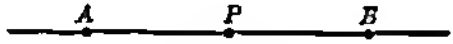


FIG. 11

is the set of points P for which the condition 'A precedes P and P precedes B' is true, along with the points A and B. The set of points $\{P \mid A$ precedes P and P precedes $B\}$ is called an open segment. An open segment from A to B is the segment from A to B exclusive of its end points. With each open segment, as with a segment, we associate a number called its length.

With these ideas in mind we might ask under what conditions can we state that 'P is close to Q'? Our physical conception of closeness might suggest that P is close to Q if the distance between P and Q is small. But what is small? A distance less than one unit? Less than one-half unit? Less than one-tenth unit? In making our definition we make an arbitrary choice of one of these lengths and work with it. One way in which we can define 'P is close to Q' is by the condition that the distance from P to Q is less than one unit. An equivalent way of stating this definition is 'P and Q belong to an open segment of length one' (Fig. 12). For a point P as shown in Figure 13 we can determine the

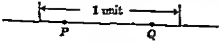


FIG. 12

FIG. 13. Open segment of length one.

set of points, Q, for which it is true that 'P is close to Q'. Locate a point A (Fig. 14) such that A precedes P and the length of the segment from

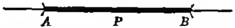


FIG. 14

A to P is one. Locate a point B such that P precedes B and the length of the segment from P to B is one. For each point Q belonging to the open segment from A to B it is true that P is close to Q.

We might inquire about the meaning of the statement 'John's house is close to Bill's house'. We would like to know first under what conditions it is true for any pair of individuals, A and B, that A's house is close to B's house. Begin by considering pairs of individuals for which the condition 'A's house and B's house are on the same street' is true. Now add the condition 'there are less than ten houses between A's house and B's house'. In this manner we can list the set of pairs of individuals for whom it is true that A's house and B's house are on the same street and there are less than ten houses between A's house and B's house. This is the same as the set of pairs of individuals for whom it is true that A's house is close to B's house by definition.

Take a picture from a magazine, locate twenty points on this picture and label them $P_1, P_2, \cdots, P_{20}$. Set a compass to measure a unit of length. Using the definition, P is close to Q if and only if there is an open segment of length one which contains P and Q. We can thus list the set of pairs of points (P, Q) for which it is true that 'P is close to Q'.

An example like this provides a setting in which we can (1) point out that even our notion of closeness or proximity yields to a precise definition, (2) consider equivalent conditions in the sense that they determine the same sets, (3) help prepare the student for future study of mathematics, particularly with regard to limits and calculus, and (4) illustrate that a closeness relation, as we have defined it, is different from an equivalence relation, such as congruence, in the following sense. Three conditions for the congruence of line segments are: (1) $\overline{AB}$ is congruent to $\overline{AB}$, (2) if $\overline{AB}$ is congruent to $\overline{CD}$, then $\overline{CD}$ is congruent to $\overline{AB}$, and (3) if $\overline{AB}$ is congruent to $\overline{CD}$ and $\overline{CD}$ is congruent to $\overline{EF}$, then $\overline{AB}$ is congruent to $\overline{EF}$. We say that the congruence relation is (1) reflexive, (2) symmetric, and (3) transitive. Although closeness is reflexive and symmetric, it is not transitive. *Things close to the same thing may not be close to each other.*

By providing experiences for the student with conditions in contexts similar to those of the preceding paragraphs, we can help him extend

earlier notions of approximation. We can help him to see that relations such as *is close to* and *is approximately* can be treated mathematically. These contexts also serve to illustrate the continued, although more formal, use of an order relation in the definitions of 'is close to' or 'is approximately'. Even more important, as we find corresponding conditions in different contexts we begin to look for a mathematical structure or structures common to these contexts.

In the high school mathematics program we should make use of every opportunity to exploit the analogous concepts of geometry and algebra. One of these is 'closeness' for pairs of points and 'approximativeness' for pairs of numbers. Each of these is defined in terms of an order relation. Making use of a one-to-one correspondence relating points belonging to a line and the set of real numbers we can develop the analogy of geometric and numerical approximation.

We illustrate the pairing of points belonging to a line and numbers in the usual manner. The pairing illustrated in the drawing (Fig. 15)

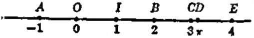


FIG. 15

can also be given in the form $B{:}2$; the point B is paired with the number 2. Similarly for $A{:}-1$, $O{:}0$, $I{:}1$, $C{:}3$, $D{:}\pi$, and $E{:}4$. We assume that if we pair a point A with a number a and a point B with a number b, the order of points belonging to the line corresponds to that of numbers.

A precedes B if and only if $a \approx b$. We define the length of the segment from A to B, provided A precedes B, $A{:}a$, and $B{:}b$ as $b - a$. We identify the open segment (Fig. 16), $\{P \mid A$ precedes P and P precedes $B\}$

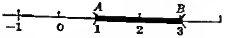


FIG. 16

where $A{:}1$, $B{:}3$, and $P{:}p$ with the set of numbers $\{p \mid 1 < p < 3\}$. Correspondingly we call the set $\{p \mid 1 < p < 3\}$ an open interval of numbers and define its length as $3 - 1 = 2$.

On page 216 we gave the definition 'P is close to Q if and only if P and Q belong to an open segment of length one'. Again, correspondingly we will define for numbers p and q, p is approximately q if and only if there is an open interval of length one which contains p and q. With this definition we can verify as either true or false each of the statements: 1.7 is approximately 2, 3 is approximately 5, π is approximately 3, $\sqrt{2}$ is approximately 1. '1.7 is approximately 2' is true since the

open interval $\{p \mid 1.5 < p < 2.5\}$ is of length one and contains 1.7 and 2. '3 is approximately 5' is false. We can verify this by showing that the assumption that there is an open interval of length one containing 3 and 5 leads to a contradiction. If the open interval $\{p \mid a < p < b\}$ is of length one and contains 3 and 5, then

$$b - a = 1, \quad a < 3 < b, \quad a < 5 < b.$$

If $a < 3 < b$, then $-b < -3 < -a$. If $5 < b$ and $-3 < -a$, then $5 - 3 < b - a$. But since $b - a = 1$ we have $5 - 3 < 1$. Consequently we have the contradiction, $2 < 1$ and $1 \nless 2$.

The idea that intervals of numbers correspond to line segments is useful in graphing. Just as we graph the set of solutions of equations such as $y = x$ and $y = x^2$ we can graph the set of solutions associated with approximation conditions such as $y \approx x$ and $y \approx x^2$. Drawing a graph of a relation helps us visualize the relation as a set of points in a plane. We make use of a one-to-one correspondence between the set of points called a plane and the set of pairs of numbers. This correspondence is illustrated by the usual method of graphing using rectangular coordinates. The definition of a relation as a set of pairs is interpreted geometrically as a set of points in a plane. Any collection of points in a plane such as a curve or a region is a relation. A student's concept of an approximation relation is enhanced when he views it as a strip of points in a plane.

To illustrate this we will use as the definition of an approximation relation, $x \approx y$ if and only if x and y belong to an open interval of length one. The graph of an approximation relation can be conveniently treated in the classroom by considering the condition $x \approx y$ along with the condition $x = y$. If we treat x and y as coordinates of a point, then we interpret the set of solutions of the equation $x = y$ as the set of points whose coordinates are equal. Similarly for $x \approx y$ we interpret the set of solutions as the set of points whose coordinates are approximately equal. We also think of the set of solutions of the equation $x = y$ as the set of points belonging to a straight line. Correspondingly we think of the set of solutions of the condition $x \approx y$ as the set of points which almost belong to a straight line with the equation $x = y$. We make use of our definition of $x \approx y$ to determine whether or not we can conclude that the point (x, y) almost belongs to the line with the equation $x = y$.

One way to interpret the definition "$x \approx y$ if and only if x and y belong to an open interval of length one" by a graph is as follows: Use the correspondence, X:x and Y:y. Locate a point X on the line segment as shown in Figure 17. The points close to X are in the open segment

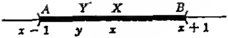


FIG. 17

from A to B. A precedes X and A is one unit below X. A corresponds to $x - 1$. X precedes B and X is one unit below B. B corresponds to $x + 1$. If Y is close to X then Y is between A and B. The corresponding condition on the coordinates x and y is $x - 1 < y$ and $y < x + 1$. For the x and y coordinates of a point in a plane this can be interpreted as stating that the y-coordinate is between one less than the x-coordinate and one greater than the x-coordinate. For example, if the x-coordinate is 3, then the y-coordinate is between 2 and 4. (3, 2.1), (3, 2.9), and (3, 3.5) are points whose coordinates satisfy the condition $x \approx y$.

To graph the set of points whose coordinates satisfy the condition $x \approx y$ we graph the set of points whose coordinates satisfy the condition, $x - 1 < y$ and $y < x + 1$. First we can graph the set of points whose coordinates satisfy the condition $x - 1 < y$ (Fig. 18). We can

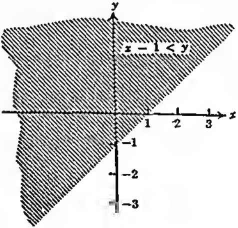


FIG. 18

also graph the set of points whose coordinates satisfy the condition $y < x + 1$ (Fig. 19). The set of points whose coordinates satisfy both of these conditions is illustrated in Figure 20 as the intersection of the regions shown in Figures 18 and 19. The points whose coordinates satisfy $x \approx y$ are those shown in Figure 20. In this figure we see these points as the points between the lines with the equations $y = x - 1$ and $y = x + 1$.

We can now compare the set of solutions of the conditions, $y = x$

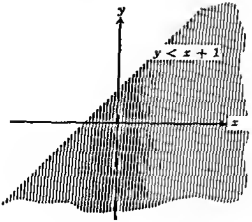


FIG. 19

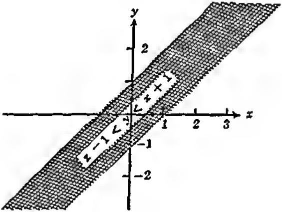


FIG. 20

and $y \approx x$. The set of solutions of the equation $y = x$ is pictured as a straight line. The set of solutions of the approximation condition $x \approx y$, is the strip of points in the plane between the lines with the equations $y = x - 1$ and $y = x + 1$.

The graph of the equation $x = 2$ in a plane is a straight line parallel to the y-axis. The graph of the approximation condition, $x \approx 2$, with the definition of approximation we are using at the moment, is the set of points between the lines with the equations $x = 1$ and $x = 3$ (Fig. 21).

As we consider the graphs of the sets of solutions of equations we can also consider the graphs of the corresponding approximation conditions. Several of these are illustrated in the Figures 22, 23, and 24.

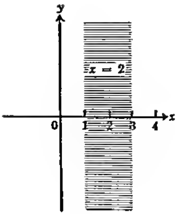


FIG. 21

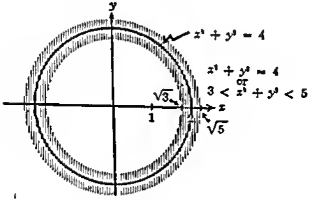


FIG. 22

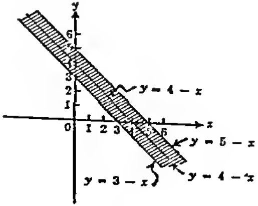


FIG. 23

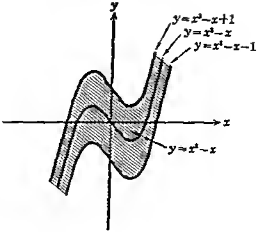


FIG. 24

Approximation in Science. A concept of approximation which develops as a student considers a measuring process, arithmetic calculations, and mathematical conditions to define an approximation relation can be useful to a science student. He needs help in learning how to formulate mathematical conditions appropriate to a variety of problem situations. A source of problem material with which a student can practice relating mathematical conditions to physical contexts is his own laboratory manual for a science course. The teacher can also use these manuals in addition to the text book to obtain illustrations of the types of situations in which a concept of approximation is used.

Common types of situations in which a concept of approximation is involved are: (1) the use of mathematical conditions which involve irrational numbers or rational numbers with no finite decimal representation, (2) the use of mathematical tables, tables of physical constants, or a computing device such as the slide rule, and (3) the use of drawings, graphs, and data to suggest mathematical conditions which define relations and functions. As a student studies algebra and geometry we can help him recognize that a term such as 'approximately' which is used rather loosely in ordinary conversation can be used precisely in mathematics and science by formulating mathematical conditions to define it. If a definition is given and never used it seems unlikely that a student will view it as important or learn it. Once an approximation relation has been defined the situations listed above provide opportunities to make use of the definition.

In a problem which involves the area of a circle whose radius is 3 inches we use the conditions '$A = \pi r^2$' and '$r = 3$'. We conclude that

$A = 9\pi$. However, 9π is irrational and may be related to some rational number by approximation. Depending upon the definition of approximation used, 9π may be related to 27, 27.9, or 28.16. Similarly from a problem which involves a right triangle and the conditions $a^2 + b^2 = c^2$, $a = 3$, $b = 2$, and $c > 0$, we conclude that $c = \sqrt{13}$. The number $\sqrt{13}$ can be related to 4, 3.6, or 3.60555. Again, the relation used is an approximation relation. $\sqrt{13} \approx 4$, $\sqrt{13} \approx 3.6$, $\sqrt{13} \approx 3.60555$. The rational number to which $\sqrt{13}$ may be related by approximation will depend upon the definition of "$\approx$" used. If $\sqrt{13} \approx x$ means $x - 1 < \sqrt{13} < x + 1$, then all of the statements $\sqrt{13} \approx 4$, $\sqrt{13} \approx 3.6$, $\sqrt{13} \approx 3.60555$ are true. However, if $\sqrt{13} \approx x$ means $x - .05 < \sqrt{13} < x + .05$, then $\sqrt{13} \approx 4$ is false, but $\sqrt{13} \approx 3.6$ and $\sqrt{13} \approx 3.60555$ are true.

In using a slide rule to compute the product (34)(19), since the numbers 34 and 19 are represented by lengths on the rule, we can only state that (34)(19) is approximately some number. Suppose it is stated that $(34)(19) \approx 640$ and that this means that the first two digits in a decimal representation of (34)(19) are 6 and 4, then the statement that $(34)(19) \approx 640$ is correct. On the other hand if $(34)(19) \approx x$ means that $x - 5 < (34)(19) < x + 5$, then the statement that $(34)(19) \approx 640$ is false, but $(34)(19) \approx 650$ is true.

In dealing with an equation such as $y = 2x - 1$, where x and y belong to the set of real numbers, we picture a set of solutions of $y = 2x - 1$ by listing a finite number of solutions, $(0, -1)$, $(1, 1)$, $(2, 3)$ $(3, 5)$ (Fig. 25), and then drawing a line through these points (Fig. 26)

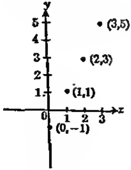


FIG. 25

Suppose, however, we have a finite set of pairs of numbers, such as $(0, -.9)$, $(1, 1.3)$, $(2, 3.1)$, $(3, 4.6)$. We can plot each pair of numbers as a point. Figure 27 indicates that we might say that these points are approximately on a line. If we interpret 'line' as the set of solutions

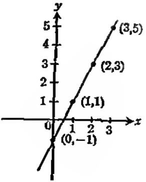


FIG. 26

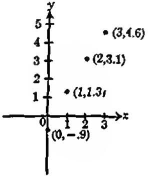


FIG. 27

of an equation $y = mx + b$, where m, b, x, and y belong to the set of real numbers, then the statement that the points $(0, -.9)$, $(1, 1.3)$, $(2, 3.1)$, and $(3, 4.6)$ are *approximately on a line* may mean that the coordinates of these points satisfy an approximation condition, $y \approx mx + b$. Although statistical conditions concerning least squares may be used to draw conclusions about the best values for m and b, we can also formulate conditions involving m and b by assuming that $(0, -.9)$, $(1, 1.3)$, $(2, 3.1)$, and $(3, 4.6)$ satisfy $y \approx mx + b$. We define $y \approx mx + b$ is defined by $y - 1 < mx - b < y + 1$. From this assumption and definition we conclude that:

$$-.9 - 1 < b < -.9 + 1,$$

$$1.3 - 1 < m + b < 1.3 + 1,$$

$$3.1 - 1 < 2m + b < 3.1 + 1,$$

$$4.6 - 1 < 3m + b < 4.6 + 1.$$

If $m = 2$ and $b = -1$, then each of these conditions is satisfied. This is sufficient to conclude that (0, −.9), (1, 1.3), (2, 3.1), (3, 4.6) satisfy the approximation condition $y \approx 2x - 1$. When the set of solutions of this condition is pictured as a strip of points in a plane, we see that the points (0, −.9), (1, 1.3), (2, 3.1), (3, 4.6) belong to the strip bounded by the lines with the equations $y = 2x - 2$ and $y = 2x$ (Fig. 28).

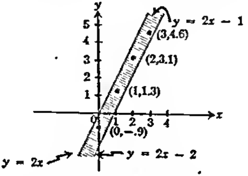


FIG. 28

Similarly when we consider sets of solutions of equations and their graphs, such as $y = ax^n$, $y = ab^{kx}$, where a, b, n, k, y, and x belong to the set of real numbers, we can also examine the sets of solutions of approximation conditions $y \approx ax^n$, $y \approx ab^{kx}$ and their graphs (Figs. 29 and 30). These ideas in a mathematics class should be useful to a

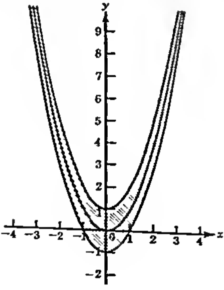


FIG. 29

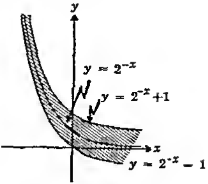


FIG. 30

science student when he collects data in the form of pairs of numbers. These pairs of numbers are in turn represented graphically with the idea that this data provides evidence that a relation is approximately linear, polynomial, or exponential.

This suggests that a common and important meeting place for the ideas of function, relation, and approximation is the determination of empirical formulas. These are formulas defining functions which determine pairs of numbers that are approximately equal to the pairs recorded as the results of measurements or observations of a statistical nature. The determination of linear empirical formulas requires the testing and recognition that a set of pairs of numbers representing observations closely approximate pairs satisfying an equation of the form $y = mx + b$, and then the determination of appropriate values for m and b. Both the testing and the determination of values for the constants interweave simple ideas of graphs, slope, intercepts, and simultaneous equations. Such work is important, useful, within the range of algebra and geometry students, and displays significant mathematical principles and relationships. At more advanced levels the same general processes of (1) testing for the type of formula to be used and (2) determining the constants for relations of the form $y = ax^n$, $y = ab^{kx}$, $y = ax^2 + bx + c$ have all the values noted above and in addition give meaningful uses of and practice with logarithms, logarithmic and semi-logarithmic graph paper, parabolic curves, and finite differences. We have not the space here to outline the processes and detail the pedagogical values implicit in teaching them in the secondary schools, but such discussions are available in many places and we recommend them to our readers.

The early recognition of order relations, greater than and less than, on sets of numbers and the use of these relations in making estimates provide experiences from which a child can abstract. This abstraction

first takes the form of the symbolic expression of '7 is less than 13' as '$7 < 13$'. Later he considers the condition '$x < 13$' where x is a natural number and recognizes that '$x < 13$' is true of some natural numbers and false of others. This is a natural setting in which he can focus attention on the set of numbers for which '$x < 13$' is true. Similarly in measuring he learns that the length of a line segment is approximately 6 inches. He learns that this may mean the length is between 5.5 inches and 6.5 inches or that the length is between 5.75 inches and 6.25 inches depending upon the unit of measure. This first acquaintance with an approximation relation serves as a background from which the conditions '$l \approx 6$' and '$5.5 < l < 6.5$' are abstracted. In this manner the physical notion of quantity gives way to the consideration of mathematical conditions and their relatedness in deductive schemes. Here, for example, the conditions are related by definition as $5.5 < l < 6.5$ is used to define the condition $l \approx 6$. This process of abstraction continues as the student is encouraged to consider any set, S, with an order relation, '$\leq$', read 'less than or equal to', defined by the properties:

(1) $a \leq a$ (reflexive)

(2) if $a \leq b$ and $b \leq a$, then $a = b$ (antisymmetric)

(3) if $a \leq b$ and $b \leq c$, then $a \leq c$ (transitive)

The concept of approximation which a student develops in the secondary school mathematics program by considering mathematical conditions involving an order of real numbers or of points belonging to a line serves as part of his preparation for the study of limit processes and other mathematical constructions which contain order relations.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in this chapter.

6

Probability

DAVID A. PAGE

Applications of the theory of probability are so numerous today that it would take an entire chapter to begin listing them. Probability and its technological brother, statistics, are used by most major industries in this country to keep check on the quality of the articles produced. Physicists develop new theories of atomic phenomena which use probability as a fundamental tool. Military experts plan the defense of our country using the results of research in probability. Since probability theory is a mathematical idealization of certain aspects of human uncertainty, it is not surprising that it has such widespread applications—so much of human activity must take place in the face of uncertainty.

If the sentences above are paraphrased for students, they may grant that the subject of probability is indeed important, but such reverence alone will not keep them studying it long. Students do not maintain daily interest in a field because they have been convinced that it is important. Rather they are interested because they enjoy it day by day, because they find it inherently fascinating. This chapter explores some ways in which a teacher might attempt to make some of the important aspects of the theory of probability alive and interesting for students between kindergarten and grade 12.

One word of caution is needed here. Probability (or statistics combined with probability) is now almost never taught as a separate subject and is seldom even touched upon in grades K through 12. Currently there is an important debate over whether statistics and probability should be introduced as a separate course in our secondary schools.* This chapter should not be interpreted as recommending the introduction of such courses. The author believes that the subject matter of probability is primarily important at the precollege level in so far as it

* See the Report of the College Entrance Examination Board's Commission on Mathematics and the preliminary text: *Introductory Probability and Statistical Inference for Secondary Schools—An Experimental Course.* It is available from the College Board.

can be used to add interest, understanding, and pleasant calisthenics to the more fundamental parts of the curriculum.

Probability theory is often thought of as dealing with the occurrence of *random events* in nature, such as whether a tossed coin will come down *heads* or *tails* or whether when dealt a hand in a game of bridge this hand will contain 4 aces or whether three bombs are adequate to destroy a factory. In recent years the theory of probability has become a purely mathematical topic which does not deal in any way with the happenings of real events around us. Just as in studying mathematical geometry one finds that lines and points are not carbon lines on a paper or chalk lines on the blackboard but rather abstractions which are suggested by such pictures, similarly mathematical probability theory is the study of an abstract formulation of probability which is suggested by the occurrence of random events in nature. From a rather small number of postulates together with a body of knowledge derived from other areas of mathematics one proves theorems concerning probability. Teachers should know that in its advanced, modern cloak, probability theory has as much claim to being a purely mathematical doctrine as any other area of mathematics.*

It is certainly not now feasible to teach probability theory as a *pure* mathematical subject in high school or earlier. But many of the notions important in probability theory can be grasped in a clear, intuitive way by children. This article tries to show how the teacher in the first twelve grades can *sprinkle in*, here and there, bits of interesting work dealing in an informal way with the ideas of probability. The inclusion of such sprinklings has at least three purposes:

1. To familiarize the student with some of the fundamental, intuitive ideas of probability and expose him to real and imagined *experiments* which should precede a thorough and rigorous study of the subject. (Example: independent events. See page 260.)

2. To provide a novel and interesting context for *standard* mathematical ideas and thereby to deepen the students' understanding of and interest in these ideas. (Example: developing space perception by working with a 3-dimensional lattice. See page 254.)

3. To give students and teachers a new and refreshing method for, in essence, *drilling* on mathematical ideas which are already taught as part of the standard curriculum. (Example: solving linear equations with *random coefficients*. See page 268.)

---

* See Robbins, Herbert. "The Theory of Probability." *Insights Into Modern Mathematics.* Twenty-Third Yearbook. Washington, D. C.: National Council of Teachers of Mathematics, 1957. p. 336.

PROBABILITY IS PROBABILITY

What can we tell students about the meaning of *probability*? It is common to say that if you toss a coin, the probability that it will come down *heads* is ½. What exactly does this mean?

Question	*Answer*
1. Does it mean that if you toss a coin two times it must come down heads one time and tails the other?	No
2. Does it mean that if you toss a coin ten times it must come down heads five times and tails five times?	No
3. Does it mean that if you toss a coin a thousand times it *must* come down heads somewhere between 450 and 550 times?	No
4. Does it mean that if you toss a coin twenty times in a row it is impossible for it to land heads each of the twenty times?	No

Then what does it mean to say that such an event has a probability of ½? In trying to change the questions given above into correct statements about probability one might say something like:

1. If you toss a coin two times, it is *more likely* that it will land one head and one tail (without regard to which comes first) than that it will land two heads.
   But isn't this the same as saying:
   If you toss a coin two times, the *probability* is greater that it will land once heads and once tails than that it will land *two heads.*
2. If you toss a coin ten times, there is a *better chance* that it will land five heads and five tails than any other single alternative.
   But isn't this the same as saying:
   If you toss a coin ten times, the single outcome with the *highest* probability is five heads and five tails (again we are not concerned with the order in which the heads or tails occur).
3. If you toss a coin one thousand times it is *very likely* that the number of heads will be between 450 and 550.
   But isn't this the same as saying:
   If you toss a coin one thousand times the *probability is high* that the number of heads will be between 450 and 550.
4. It is most unlikely that if you toss a coin twenty times, it will land heads each of these twenty times.

But isn't this the same as saying:

The *probability is very low* that if you toss a coin twenty times, it will land heads each of these times.

Thus every time you make a correct statement of what to expect of a coin-toss, you are forced to use some expression (perhaps more disguised than those above) which contains a word such as *chance, likelihood, unlikely, almost certain, almost never,* and so on. But these expressions are essentially synonymous to an expression containing the word *probability* itself. *The explanations are circular.*

One way to tell what *probability* means is to give a collection of postulates which tell how to use the word mathematically. Since we have already indicated that a postulational approach to probability is beyond the capability of school children, what can a teacher do who wants to introduce some of the ideas of probability? The answer is: Use the word *probability* as accurately as you can, work with the ideas of probability which follow, and don't try to say what probability is. Students are quite accustomed to picking up ideas from context. Keep the context reasonably correct and they will infer correct notions.

ACTIVITIES FOR ELEMENTARY SCHOOL

Throughout the elementary grades there are many opportunities to give children a better understanding of standard topics while giving them some ideas about probability. In kindergarten and first grade, children can carry out a project of *dice-casting.* Give each student a die or let students take turns with fewer dice. The goal is to make a large number of throws and to record the results. If throwing dice is objectionable, students can spin pointers or select pieces of paper with numerals on them from a hat. When a student throws a die and *reads it,* he must look at the spots and identify the number to which the set of spots corresponds. The results are recorded on a chart such as illustrated in Table 1.

TABLE 1

1	2	3	4	5	6
///// //	///// /	////	///// /////	/////	///// /

The chart form can be one that the student has made for his own use, or it can be duplicated in advance by the teacher, or there can be a single chart on the blackboard for the entire class to use. With children

who are quite new at working with numbers and who are likely, therefore, to make many mistakes, the single chart is better. Each student should come to an *experiment table* to make his throws so that the teacher can check on his decisions concerning the results of his throws. If a nondice experiment is used, the number of columns in the chart must be adjusted to the number of possible outcomes. At the beginning it should be between 5 and 10.

The student records his result by making a tally mark in the appropriate column of the chart. (In doing this he has to *recognize* an Arabic numeral.) To make it easy to compare the number of entries in each column some definite pattern of making the marks should be agreed upon. In the example above the agreement was that after five marks have been made in a row the next mark begins the next row. Here is an ideal place to introduce the *standard system of tallying:*

~~||||~~ ~~||||~~ ||

Also, here is the place to encourage students to *invent* other methods of keeping track. Continue the *experiment* until there has been a total of at least one hundred throws recorded on the chart. With students who have used individual charts at their seats, collect all the data of the entire class onto a single chart of this type. Now there is an opportunity for counting. Count the number of tallies in each column. Also count the total number of tallies. For practice, count the number of tallies in the first three columns and the number of tallies in the last three columns, and so on. Usually somewhere *fairly near* $\frac{1}{6}$ of the total number of tallies will occur in each column. In particular *it is quite* unlikely that in one hundred total throws, any column will be *missed* completely. It is also quite unlikely if, say, 102 throws are made that exactly 17 tallies will appear in each column. Just by noticing the *random* way in which the tally marks were put down in the first place, students develop a better intuitive idea of randomness.

Ask students questions like these:

1. If we only made six throws, would it have to be that one mark landed in each column? (No)
2. If we made six throws and then made another six throws and then another six throws, and so on, would very many of these individual *six-throws* give exactly one mark in each of the six columns?

   (Answer: only a few would. If students don't agree on this conclusion or don't feel strongly about any conclusion, have them carry out the experiment.)

3. After seven throws were made, could one column still remain empty? (Yes)
4. After ten throws were made, could one column be empty? (Yes)

Continue asking questions like this going up five throws at a time, asking about fifteen throws, twenty throws, twenty-five throws, and then moving ten throws at a time until you get to, say, the following question:

5. If you made a hundred throws, would it be possible for one column to be empty? (Yes, it would be *possible* but it wouldn't happen very often.)
6. If you made a hundred throws and then another hundred throws and then another hundred throws and so on, would very many of these *hundred-throws* give you charts in which one column was completely empty? (No, *very* few of them would.)

For classes that are interested in the topic and able to go farther with it, there are many more questions that you could ask at this time, many of which would provoke dispute and need to be settled by experiment:

7. Does the location of the first, say, twenty marks in the table affect where the next mark will go? Specifically, suppose that after twenty marks are made, one column is blank; does that mean it is pretty certain that a mark will next fall in the blank column?* (No. But some students may be hard to convince.)
8. Suppose we throw ten dice once instead of one die ten times. How will the outcome be affected? (You wouldn't expect the exact same results, but the *chances* will remain the same.)
9. If you throw 4 dice over and over again about how often will they all come down sixes? (About once every 1300 times. A good answer from a student at this stage might be: much less than once every 100 times.)

Students will have no other way than experimenting to answer such questions except for the intuition of some gifted students.[1] You will see how to handle such questions mathematically later in the chapter.

The student should gain from this early work with dice or similar devices the feeling that *"the chances" of one throw landing in any one of the six columns are the same as the chances of landing in any other of the six columns.* In more formal terms (not necessary for use with students) he should realize that the six possible outcomes of a single throw are *equally likely.* His understanding of

* This question alludes to the frequently encountered (but *fallacious*) notion of the *Law of Averages*. If you toss a coin, say, four times and get all heads, on the next toss you are almost certain, by the "Law of Averages," to get a tail on the fifth try. The error in this kind of thinking will be more apparent later in the chapter when Independent Events are discussed.

more complicated ideas will depend upon his gaining first an accurate notion of *equally likely events.*

Some of the ways (*counting, tallying,* numeral recognition) in which these dice throwing and similar activities reinforce ordinary work in early arithmetic have been indicated above. You can increase the arithmetic complexity of such activity by substituting for dice regular solid figures of more than six sides with numerals on the sides.* If you use a solid which does not have pairs of parallel sides, you will need to read the side which is down instead of following the usual procedure of reading the side which is up as you do with dice.

An instructive comparison or, better yet, a preliminary to this exercise is the experiment of *flipping* a coin and recording the results in a two-column chart which has one column labeled *heads* and the other column labeled *tails.* Have students in the early grades report that they are learning *coin flipping* at your own risk! However, if you want to bring in this idea, you can use checkers or similarly shaped objects with *one* on one side and *two* on the other or *5* on one side and *10* on the other.

There are also games for children on the market which use plastic cubes labeled with letters of the alphabet rather than with the *spots* of ordinary dice.† Such cubes can be used to study spelling and probability together as will be pointed out later. In particular they can be used for the student activities suggested thus far. Ultimately, numbered cubes will be more useful.

If shop facilities are available, wooden cubes an inch or two on an edge can be labeled with letters, numerals, or spots. Or they can be coated with blackboard paint and then they can be labeled and relabeled with chalk to suit the activity at hand.‡

Many simple probability devices can be constructed and appreciated by elementary school students. They can open a book *at random* and take the last digit of the page numeral. Be careful of well worn books that tend to open at certain pages! The numerals in telephone directories

* Caution: If you use a many-sided solid which is homemade out of paper or cardboard, it may easily turn out that the distribution of weight in it is not symmetric—some joint may have a lot of glue in it—and then the side opposite the heaviest side will be favored. The sides will not give equally likely outcomes.

† One such game is Throw and Spell manufactured by the Toycraft Company, Chicago 5, Illinois. Along with 15 alphabet cubes is included a cube labeled with Arabic numerals which eliminates any stigma that may be attached to ordinary dice.

‡ The blocks must be quite accurately cubical or the various faces will not be equally likely. This suggests an appropriate variety of topics for able high school students looking for *projects.* Build rectangular solids of various shapes. Try to find a connection between the dimensions of the blocks and experimentally determined probabilities. Extend to nonrectangular solids.

can be used in many similar ways. Challenge your students to think up other systems which give equally likely outcomes. If there is some doubt about a student's system, test it by repetition. For example, if an experiment has three possible outcomes and after 100 repetitions, two alternatives have occurred 25 times each and the other alternative has occurred 50 times, you know that either a very unlikely event has occurred or else the alternatives are not equally likely. If you want to make a careful test, you are out of the range of students in the early grades. For your own information or for older students, consult the chapter on "Statistics" in this Yearbook.

ACTIVITIES FOR MIDDLE GRADES

Somewhere around grades three to five, students can carry out experiments similar to the one just described but in which they throw *two dice* at a time. In order to tell the two dice apart, use dice of different colors. Using dice that are easy to tell apart is extremely important; don't decide at the last minute to use the dice you used before if they are imperceptibly different from each other. When students were throwing only one die, it made no difference if different students had different colors of dice. Therefore, if you purchase them, it might be wise to buy differently colored dice, say, three different colors at the outset. (For obvious reasons, it is probably wise for the school to furnish the dice as *experimental equipment* and for the teacher to collect them at the end of each experiment.) Call one of the two dice the *first die* and the other the *second die*, and keep these names the same. If you are working with one red and one white die, you might agree to call the red one the *first die*, and the white one the *second die*. (If students have trouble keeping this straight, you can use a mnemonic device such as "*first*, in the morning there is a *red* sunrise and *second*, during the day it gets *white*".)

Have students make many throws, each time throwing the pair of dice. A good procedure to follow here is to have students work in pairs, one student throwing the dice and the other student recording the results. After a while students can interchange duties. Students can record their results on a chart such as illustrated in Figure 1.

The design of this chart, in particular the horizontal choice for *first die* and vertical choice for *second die*, is not accidental. It is intended as an analogue to plotting points in the first quadrant of the coordinate plane. This analogy will be more apparent when we come to the next version of this same experiment.

Notice that each time a student makes a throw of two dice, he obtains

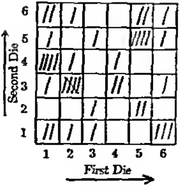


Fig. 1

two numbers, a first number and a second number. The first number is given by the first die and the second number by the second die. In technical language, the student obtains an *ordered pair* of numbers. If you explain to students that an ordered pair of numbers is a *pair* or *couple* of numbers where you know that one of the numbers is the first and the other number is the second, students could probably make more sense out of the term *ordered pair* than many of the words they are sometimes expected to be familiar with in arithmetic (for example, 'subtrahend' and 'minuend').

At this point it is worthwhile to have students invent a notation for recording the results of a single throw when they do not make use of a chart as above. For a throw in which the first die gives 3 and the second die 5, typical notations that inventive students might suggest are illustrated in Figure 2. You might point out to students that a notation

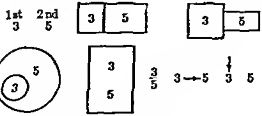


Fig. 2

which would be good for them to know about because they will use it in a very similar way later on in school is the following:

$$(3, 5).$$

A good *drill* activity which students will hardly think is drill is the following: let one team throw a pair of dice and using the notation

suggested above, (3, 5) and so on, have them record the results of each throw on a blank sheet of paper rather than a rectangular chart. Thus, such a team might produce a page which begins as illustrated in Figure 3.

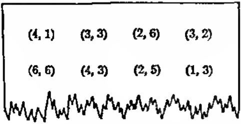


FIG. 3

Then the second team working from this *data sheet* transfers the data onto a 6-by-6 chart of dots like the one shown below. Notice that the talley marks improve the analogy between the activity of the second team and *plotting points* in the first quadrant of the coordinate plane. After recording 40 ordered pairs the chart might look like the one illustrated in Figure. 4

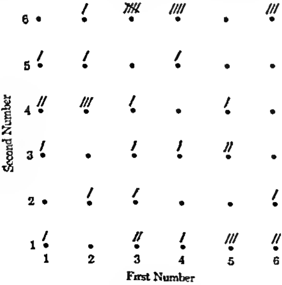


FIG. 4

You can extend this process of recording ordered pairs of numbers on a chart to bigger numbers. For example, you can give students a dittoed sheet with a *20-by-20* array of dots and another sheet which simply lists many ordered pairs of numbers constructed from the whole numbers 1 through 20. Instead of merely using a list of ordered pairs

students could obtain these ordered pairs of numbers by tossing two regular icosahedra (20-sided solids) or by spinning pointers each of which points to one of twenty different numbered regions. Also if you want to reinforce work with fractions, instead of giving them an array of dots numbered 1, 2, 3, ⋯, you can give them an array of dots numbered 1, $1\frac{1}{3}$, $1\frac{2}{3}$, 2, $2\frac{1}{3}$, ⋯. In a similar manner you can use decimals. If you have evenly spaced dots, it is probably wise at this stage to use only sequences of numbers in arithmetic progression, that is, where the difference between successive numbers is the same.

AN EARLY TASTE OF GRAPHING AND COORDINATE GEOMETRY

Before going further into the idea of probability, you now have developed the tools for giving lots of interesting and important exercises. Here is an example of a sequence of questions which you now might ask your students:

1. If Bill throws and gets (2, 5) and Tom throws and gets (5, 2), will they each make the corresponding tally mark by the same point? (Answer: no.)
2. Where will they make their tally marks? (Answer: illustrated in Figure 5.)

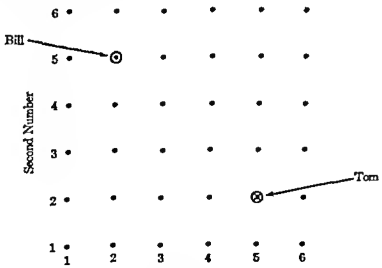


FIG. 5

3. If Bill makes marks for (2, 3) and (2, 4) and (3, 5) and (4, 6) and Tom makes marks for (3, 2) and (4, 2) and (5, 3) and (6, 4), show

where the location of each of their sets of marks would be. (Answer: illustrated in Figure 6.)

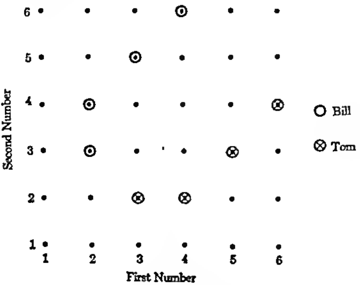


Fig. 5

4. Suppose Bill's marks were like those shown in Figure 7, and suppose that in each case Tom had the *reverse*, that is whenever Bill

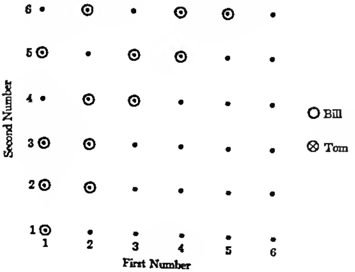


Fig. 7

had something such as (1, 5), Tom had the reverse (5, 1). What would Tom's chart look like?* (Answer: illustrated in Figure 8.)

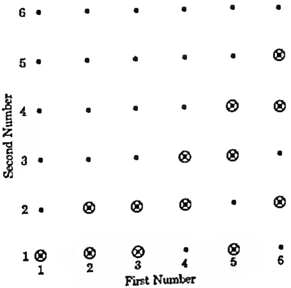


FIG. 8

Watch the students as they construct Tom's chart. Some will go to each point and find its ordered pair of numbers, interchange first and second numbers, and then find the corresponding new point. Such students are operating correctly but they haven't yet *caught on.* When they catch on, they will construct the second chart just by looking at the first one. Give more such questions to children who have not caught on. Finally give them Bill's chart as illustrated in Figure 9. What does Tom's *reverse* chart look like? By now many of them should be ready to reply almost instantly: "It's the same chart!" The reverse of Bill's chart is precisely the same chart.

When students see the easy way to do such problems, they have observed that one set of points is symmetric to the other set of points with respect to a diagonal line (Fig. 10).

Do not expect students to be able to tell you what 'symmetric' means or even to learn the word unless you do considerably more work on the topic. But such a brief exposure to the notion of symmetry will pay off when they meet it again more formally.

---

* It would also be a good idea to use the word *inverse* here from time to time in place of *reverse* in preparation for later work in mathematics. See the discussion on the inverse of a function in the chapter on "Relations and Functions" in this Yearbook.

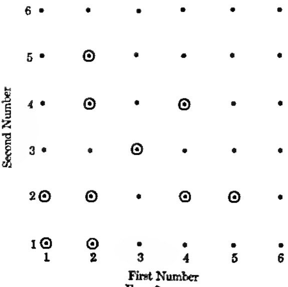


Fig. 9

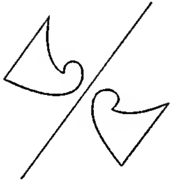


Fig. 10

Here are a few other questions which you could ask students at this stage to provide early groundwork for more formal work in graphing of inequalities later on.*

1. One time when Bill threw the dice several times it turned out that the number he got from the first die was always more than the number he got from the second die. For example, he got (5, 1), (2, 1), (6, 5), (4, 2) but never a pair such as (2, 4) or (5, 5). What can you say about how his chart looked?

* For a discussion of the graphing of inequalities and to see how this work sets the stage, see Chapter 3 of this Yearbook.

Answer: An occasional brilliant child will see immediately the *region* in which Bill's ordered pairs must fall. Most students will need to start plotting pairs with first number bigger than second number. If necessary, a student can plot all 15 such ordered pairs. Most students will catch on after plotting a few of them. For each of Bill's throws, it must have fallen somewhere within the following region of circled dots (Fig. 11).

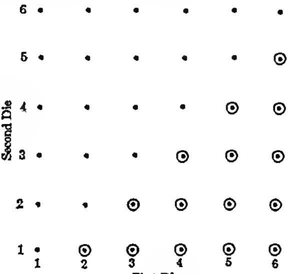


FIG. 11

2. In all of Bill's throws another time, it turned out that the number he got from the first die was never less than (but sometimes equal to) the number from the second die. What can you say about how his chart looked?

   Answer: All of the throws fell somewhere in the following region of darkened dots (Fig. 12).

For the last two questions students have been graphing inequalities using a very restricted domain of numbers. More questions of this type and related student activities will be suggested later in this chapter.

PROBABILITIES WITH TWO DICE

After students have had some practice in throwing dice and finding the corresponding point for each throw on a 6-by-6 chart (or *lattice*) of dots, ask them which dots are the *favorite* ones. That is, ask them

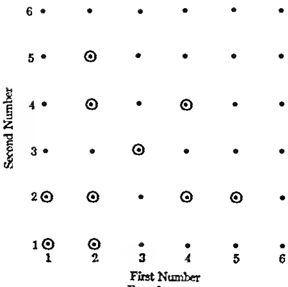


Fig. 9

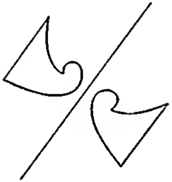


Fig. 10

Here are a few other questions which you could ask students at this stage to provide early groundwork for more formal work in graphing of inequalities later on.*

1. One time when Bill threw the dice several times it turned out that the number he got from the first die was always more than the number he got from the second die. For example, he got (5, 1), (2, 1), (6, 5), (4, 2) but never a pair such as (2, 4) or (5, 5). What can you say about how his chart looked?

* For a discussion of the graphing of inequalities and to see how this work sets the stage, see Chapter 3 of this Yearbook.

Now in the array of 36 dots draw an oval which includes two dots (Fig. 13).

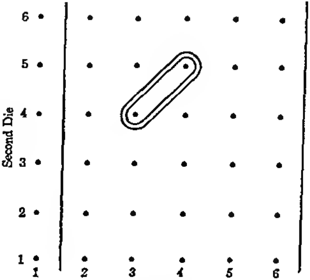


FIG. 13

Ask students what the chances are that in *one* throw of two dice, they will *get* one of the dots inside the oval. *Two chances in 36.* Now, draw a vertical line to the right of the leftmost column of dots as in Figure 13 and ask what the chances are that in one throw they will get a dot to the left of the line. *Six chances in 36.* As a bit of humor (but with a purpose), draw a line at the right of the rightmost column and ask what the chances are of getting a dot to the right of that line. Students will laugh, but press for an answer to the question. In time someone will see that he can give you a perfectly good answer using the same language pattern as was used for the other answer: *no chance in 36.* Ask many more questions of this type singling out various regions of dots. In some cases choose regions in which the dots are widely separated such as those illustrated in Figure 14.

Answer: for the solid loop 3 chances in 36, and for the dashed loop 9 chances in 36.

Eventually, encircle the entire array of 36 points and ask for the chances that in one throw they will get a dot within the loop. *36 chances out of 36.*

The next step is to recognize that expressions such as *two chances in 36* and *one chance in 18* are synonymous. With students who are familiar with fractions, this will be an easy step. For students who do not know

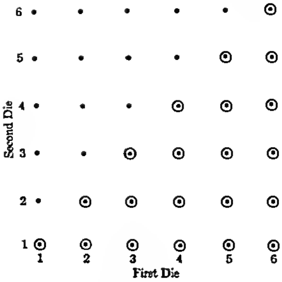


Fig. 12

if there are some dots on which one has a *better chance* of landing than on the others. When presented with such questions, a few students may claim that the central dots are more likely, and an occasional student may claim that the dots which involve sixes are more likely. The first misconception probably comes from general intuitive feeling that the *middle is better than the edge.* The second misconception comes from a *misapplication* of *correct* knowledge about dice, e.g., *getting a 6 is more likely than getting a 2.* The student who thinks a 6 is more likely than a 2 has forgotten for the moment that his ideas are correct when the 2 and the 6 are the *sum* of the two numbers from two dice (and he is ahead of the story as it is developed here). As yet we are talking about the two numbers separately without adding them.

The majority of students will conclude correctly that there is as good a chance of landing on any dot as there is of landing on any other dot. But how many different dots are there as choices? Thirty-six. Then, what are the chances of landing, say, on the upper left-hand dot? *One chance in 36.* What are the chances of landing on the dot at the lower right-hand corner? *One chance in 36.* What are the chances of landing on the dot corresponding to (3, 2)? *Still one in 36.* The language *one chance in 36* comes easily to students and they should be encouraged to use it a while before making the next step: saying, *one thirty-sixth.* In fact, students who have not studied fractions sufficiently will need to continue to follow the *so many chances in so many* pattern.

of division.* In any case, students should determine that there are 18 ovals in the figure and that one's chances of landing in one oval are as good as those of landing in any other oval. Thus, the chances of landing in the first oval discussed are *1 out of 18*. Similar grouping procedures should be carried out until students can translate rather freely from *10 out of 36* to *5 out of 18* or from *18 out of 36* to *1 out of 2*. Students who know about fractions will see that an expression such as *18 out of 36* resembles a fraction such as $\frac{18}{36}$ in that in each case one is thinking about two whole numbers 18 and 36 (an ordered pair of integers). Also an expression such as *18 out of 36* is simplified in a way similar to the simplification of fractions to *1 out of 2*. Students who are ready to work with fractions should now shift from, for example, the expression *the chances are 3 out of 5* to the expression *the chances are* $\frac{3}{5}$. Then it is an easy step to say *the probability is* $\frac{3}{5}$. The rest of this chapter will follow this pattern of wording. Teachers who do not want to bring in fractions will need to translate back to the less technical language.

Next have students write beside each of the 36 dots a numeral which gives the *sum* of the corresponding first number and second number† (Fig. 16).

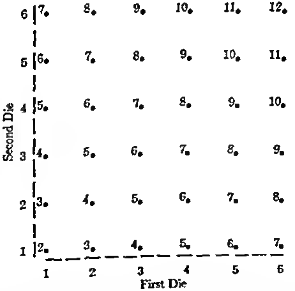


Fig. 16

* Experienced teachers will recognize that the grade level under discussion jumps about rather erratically. Such discontinuities seem to be necessary in a single chapter which deals with 13 grades. It is hoped that each teacher can extract a consistent story for the level with which he is concerned.

† Here is the place where it becomes important that each die gives a number rather than say a letter

Ask students questions such as these:

Find all the dots where the sum is, say 6. Put a loop around them. How many are there? (5)

What is the probability that if you throw two dice, the sum will be 6? ($5/36$)

What is the probability that the sum will be 2? ($1/36$)

What is the probability that the sum will be 6 or less? ($15/36$)

What individual sum is most likely, that is, what individual sum has the highest probability? (7, because there are more dots corresponding to 7 than any other sum.)

What is the probability that the sum will not be 12? (There are 35 dots corresponding to *not getting a 12* so the probability is $35/36$.)

What is the probability of getting a sum which is *7 or 8*? There are 11 points corresponding to *getting 7 or 8*. Notice here the language training in a precise use of the word 'or'. One has obtained *7 or 8* if he has obtained 7. Also, one has obtained *7 or 8* if he has obtained 8. Thus, the probability of *getting a 7 or 8* is $11/36$.*

What is the probability of getting in one throw 7 *and* 8? Since it is impossible in one throw to obtain a sum which is *both* 7 and 8, the probability of *getting a 7 and an 8* is 0. Here the student is presented with a sharp contrast between the words 'and' and 'or'. What is the probability of *getting less than 7 as a sum and getting 2 as a sum*? The event *getting less than 7 and getting 2* occurs only when you get 2. Consequently, the probability of *getting less than 7 and getting 2* is $1/36$.

What is the probability of *getting less than 7 or getting 2*? This event occurs any time the result of the toss has a sum less than 7. Thus, the probability of *getting less than 7 or getting 2* is $15/36$. Note that while it is good practice with fractions to reduce '$15/36$' to '$5/12$', when working with probabilities it is usually better not to reduce fractions but to leave them all with the *highest denominator* since this is the form in which you add, subtract, and compare them.

Now you are in a position to investigate one of the fundamental *facts* about probability. The probability of *getting a sum of 7* is $6/36$. The probability of *getting a sum of 6* is $5/36$. The probability of *getting a sum of 6 or 7* is $5/36 + 6/36$ or $11/36$. Does it always work that way?

* The word 'or' is used in two ways in common language. Sometimes '*A* or *B*' means '*A* or *B* but not both' and sometimes it means '*A* or *B* including the possibility of both'. In probability work the second meaning is usually intended. In the example just given it does not matter which way you interpret 'or' because it is certainly impossible to get both a 7 and an 8 as a sum in one shake. See the chapter on "Proof" in this Yearbook for a further discussion of the two uses of 'or'.

No. For example, the probability of getting a sum which is greater than 3 is $^{33}/_{36}$, and the probability of getting a sum which is less than 11 is $^{33}/_{36}$. Is the probability of getting a sum which is greater than 3 *or* less than 11 equal to $^{33}/_{36} + {}^{33}/_{36}$? No. (It is only $^{36}/_{36}$; any number is greater than 3 or less than 11.) By now you see that:

A probability is a number which must be greater than or equal to zero and less than or equal to 1.

No event has probability $^{33}/_{36} + {}^{33}/_{36}$. For what events can you find the probability of *one event or a second event* by adding the separate probabilities? If you refer to the six-by-six lattice of points for two dice, it is easy to decide. The probability of the event corresponding to *getting a point in the solid oval or the dashed oval*, as in Figure 17, can be found by adding the probability of *getting in the solid oval* and the probability of *getting in the dashed oval*. But (Fig. 18) the probability of

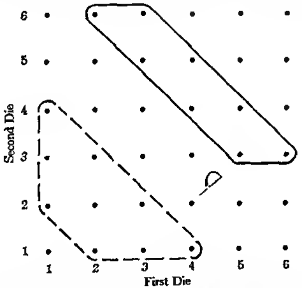


FIG. 17. Probability of getting in the solid loop: $^{8}/_{36}$; probability of getting in the dashed loop: $^{9}/_{36}$; probability of getting in the dashed loop or solid loop: $^{17}/_{36}$.

getting a point either in the dashed *oval* or in the solid *oval* is not the sum of the separate probabilities. Looking at charts it is easy to see that you can *add probabilities* to obtain the probability of an *or-event if* the two loops involved do not have any points in common. What can you say about the physical events themselves if their corresponding loops do not have any *points in common?* If there is a point in both loops, then there is a way in which both events can occur simultaneously. In

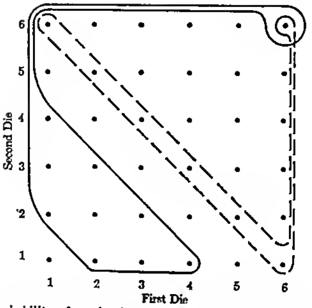


FIG. 18. Probability of getting in dashed loop: $7/36$; probability of getting in solid loop: $10/36$; probability of getting in dashed or solid loop: $16/36$.

the second illustration above, if the sum is 12, both of the events, the solid loop event and the dashed loop event, have occurred.

Events which cannot occur simultaneously are called *mutually exclusive* events. In the immediately preceding figure, the event corresponding to the dashed loop and the event corresponding to the solid loop are *not* mutually exclusive. However, the event *getting a 5 or a 6* and the event *getting a 9* are mutually exclusive. If one of them happens, the other one cannot happen.

From the point of view of probability there is a more precise way to say *two events cannot occur simultaneously.* Suppose to abbreviate we call one event 'A' and another event 'B'. Then if A and B cannot occur simultaneously, the probability of (A *and* B) is zero. That is, if there are no points in the *overlap* of A and B then the probability of landing in the overlap is zero. To summarize then, we can say that for an event A and an event B, if the probability of (A and B) is zero, then the probability of (A or B) is the sum of the probability of A and the probability of B. In order to save space and make reading easier we shall abbreviate *the probability of A and B* as "$Pr(A$ *and* $B)$. Using this abbreviation a fundamental principle of probability becomes:

(I) For events A and B, if $Pr(A \text{ and } B) = 0$, then
$$Pr(A \text{ or } B) = Pr(A) + Pr(B).$$

Now can we generalize this principle so that in its new form it applies to sets of dots that do overlap? The trouble with sets of dots that do overlap is that when you add the corresponding two probabilities you count the points in the overlap twice. The set of points in the overlap corresponds to the event *(A and B)*. Thus to get rid of the effect of counting the overlap twice you need to subtract it once. That is, you

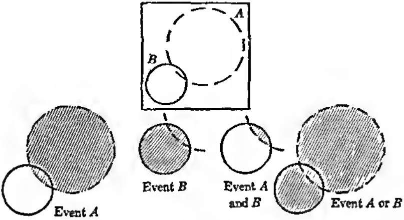


FIG. 19

need to subtract $Pr(A \text{ and } B)$ (Fig. 10). Thus the new generalization becomes:

(II) For any events A and B, $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \text{ and } B)$.

Note that generalization (I) is a special case of generalization (II) where $Pr(A \text{ and } B) = 0$.

Here is an example of generalization (II). Suppose that

event A is getting a 2 or a 7

and

event B is getting an 11 or a 7.

Then

$$Pr(A) = \tfrac{7}{36}$$

and

$$Pr(B) = \tfrac{8}{36}$$

and

$$Pr(A \text{ and } B) = \tfrac{6}{36}.$$

A is a subset of B. It will be left for teachers who are working with set theory to translate the rest of the chapter into set theoretic terminology.*

PROBABILITY DIAGRAMS IN MORE THAN TWO DIMENSIONS

Next, we want to extend our treatment of probability to situations where more than two dice or other things are used. As a first step in this extension, students should recognize that the array of 36 dots in the previous discussion was a square array only as a convenience. For example, we would cut the square array into *strips* of six dots each, such as illustrated in Figure 21. Then the set of six encircled points above

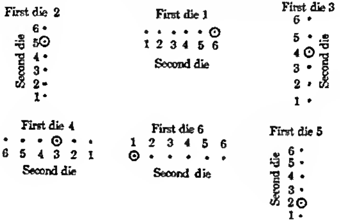


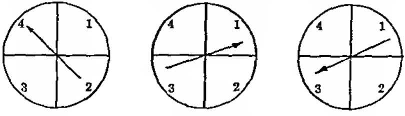
FIG. 21

corresponds to the event *getting a sum of 7*. It is more difficult to work with such a *scattered* diagram than it is with a 6-by-6 square but there is no fundamental difference. With patience we could do anything with the scattered diagram that we can do with the square array. In fact, we could take apart each row of dots until we merely had 36 separate dots splattered miscellaneously. As long as we know which dots correspond to which outcomes with the dice, we could work with any such diagram. You will see that when more than two objects are being considered, it is helpful to work with diagrams which have been partially cut apart.

Suppose you have three pointers which you spin and each pointer

* For a more thorough treatment of the theory of sets see Chapter 3 of this Yearbook and Chapter III of the Twenty-Third Yearbook.

gives you one of the numbers 1, 2, 3, or 4 with equal probability (Fig. 22). When you spin all three pointers, they give you a first number, a



First Pointer Second Pointer Third Pointer

Fig. 22

second number, and a third number. That is, they give you an *ordered triple* of numbers. A convenient way to record, for example, that the first pointer gave a 4 and the second pointer gave a 1 while the third pointer gave a 3 (as shown in the figure) is to write, similarly to the previous convention:

$$(4, 1, 3).$$

Now what is the probability that when you spin the three pointers, the result will be (1, 4, 2)? By now, your students should see that there are $4 \times 4 \times 4$, or 64 possible ordered triples as outcomes. If some of them have difficulty seeing that there are 64 possibilities, have them start to list all possibilities. Most students will see a system before they finish the list of 64 ordered triples. As before, if we assume that the pointers have good bearings and if the field behind each of them is symmetrically divided into four regions, then these 64 ordered triples are equally likely outcomes. Therefore, the probability of a particular outcome, that is, of a particular ordered triple such as (1, 4, 2) is $1/64$. The probability of any other ordered triple is also $1/64$. What kind of a diagram of dots shall we make to illustrate this situation? Judging from the previous examples, it ought to have 64 dots. (Students will easily see that if there were only two pointers instead of three, a square array of 4-by-4 points would do the job.) It should not be difficult for your students to see that in this case what we need is a "square array" of 4-by-4-by-4 dots; in other words, we need a cubical array which is 4-by-4-by-4. Drawn with only the outside dots showing, such an array is pictured in Figure 23. It is difficult to work with such a picture because *inside dots* are hard to locate. Elementary school children should build a 3-dimensional model out of balsa wood or Tinker Toys and devise some way of marking which dot is under discussion. For example, they could tally by hanging a small piece of wire over the spot corresponding to a particular *spin* of the three pointers. After establishing such an arrangement you can repeat the

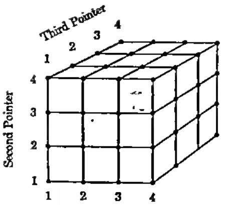


FIG. 23

same kind of exercises that you did with a 6-by-6 array of dots. For example:

Where are all the points corresponding to *second pointer gets a 1*?
Where are all the points where the sum of the three numbers is 4?
What is the probability of getting a sum 4?
What is the probability that the sum of the three numbers will be less than 6?*

Notice the training in space perception and preparation for harder solid geometry problems here. By doing more work of this kind you increase the students' intuitive background in work with 3 dimensions.

For older students and after younger students have worked with a 3-dimensional model until they are familiar with it, it is important that they cope with the problem of dealing with such a three-dimensional figure on two-dimensional paper. To do this they can cut the 3-dimensional figure into *slices*. In order that these slices resemble the previous array of 6-by-6 dots, let the first slice be the plane of 16 dots nearest to the paper (as you see the picture, assuming that the other dots are *behind* the paper). The first slice we take off of the cube looks like Figure 24. As we cut off successive slices, we get figures like the one above but with different third pointer numbers. Since there are four such slices, the three-dimensional figure can be pictured as in Figure 25. Students should see easily that there are 64 points, and that to each *spin* of the three pointers there corresponds exactly one point, and conversely to

* This is a fairly tough question and students may need to locate one by one all of the points for which the sum of the three numbers is either 3, 4, or 5, that is, less than 6. Notice that students are graphing an inequality in 3 dimensions as they answer this question.

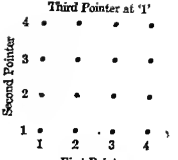


Fig. 24

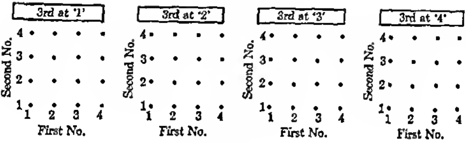


Fig. 25

each of the 64 points, there corresponds exactly one *spin* of the three pointers. Previously we *sliced up* a 2-dimensional 6-by-6 figure into six miscellaneously placed 1-by-6 strips. Now we have *sliced up* a 3-dimensional 4-by-4-by-4 figure into four 2-dimensional 4-by-4 figures placed so that they are convenient to use in order to keep track of *spins*. Now you are in a position to ask the following kinds of questions:

1. What is the probability of getting a 2 on the first pointer, a 3 on the second pointer, and a 4 on the third pointer? (*Answer:* $\frac{1}{64}$. *There is only one of the 64 points which corresponds to* (2, 3, 4).)
2. What is the probability of getting a 1 on any one of the pointers, a 2 on any other pointer, and a 3 on the other pointer? (Answer: $\frac{6}{64}$. Since the question does not specify which pointer gives which number, any of the following outcomes will do: (1, 2, 3), (1, 3, 2), (3, 1, 2), (3, 2, 1), (2, 1, 3), (2, 3, 1).)
3. What is the probability that the sum of the numbers given by the three pointers is 4? (Answer: $\frac{3}{64}$. *The following ordered triples give a sum of 4:* (2, 1, 1), (1, 2, 1), (1, 1, 2).)
4. What is the probability that the sum of the three numbers is 12? (Answer: $\frac{1}{64}$. The only way that the sum can be 12 is for each pointer to give 4. There is one point corresponding to (4, 4, 4).)

5. What is the probability that the sum of the numbers given by the three pointers is 15? (Answer: 0.)
6. What is the probability that the sum of the three numbers given by the pointers is 6? (Answer: The possible ordered triples are

$$(1, 1, 4) \quad (2, 1, 3) \quad (3, 1, 2) \quad (4, 1, 1)$$
$$(1, 2, 3) \quad (2, 2, 2) \quad (3, 2, 1)$$
$$(1, 3, 2) \quad (2, 3, 1)$$
$$(1, 4, 1)$$

so the probability is $^{10}/_{64}$. The students should be encouraged to notice that these points are located in a regular fashion (Fig. 26).)

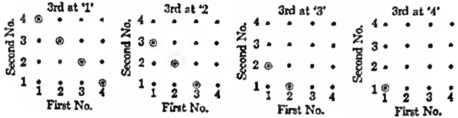


FIG. 26

In the 3-dimensional cube these points are a *diagonal slice.* (Teachers should notice the preparation for graphing the equation '$x + y + z = 6$' in 3-dimensional coordinate geometry.) More elementary students who have constructed a cube of 64 points can locate these 10 points and see directly that they comprise a *diagonal slice.*

7. What is the probability that the sum of the three numbers given by the pointers is 7? (Answer: $^{12}/_{64}$. By now students should be devising systematic ways to determine how many ordered triples give 7 as a sum. For example, there are 6 triples involving the numbers 1, 2, and 4. There are 3 triples involving the numbers 1, 3, and 3 and there are 3 triples using the numbers 2, 2, and 3. These 12 are all the triples that give a sum 7. How do you know?)
8. What is the probability that the sum given by the three pointers is 9? (Answer: $^{10}/_{64}$.)
9. What sum has the highest probability? (Answer: The sums 7 and 8 each have probability $^{12}/_{64}$ which is the maximum probability. Notice that if you consider all the possible sums,

$$3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

there is no one *center number* but 7 and 8 are on either side of the center.)

10. What sum has the lowest probability? (Answer: The sums 3 and 12 each have probability $\frac{1}{64}$ which is the minimum probability.)
11. What is the probability that the number given by the first pointer minus the number given by the second pointer plus the number given by the third pointer is zero? (Answer: If (x, y, z) is an ordered triple, then we are looking for points where $x - y + z$ is 0. For the points we want, it must be that $x + z = y$. (Elementary school students or others who are not familiar with algebra can just look for the ordered triples which work):

$$\begin{array}{ccccc} & (1,3,2) & (1,4,3) & \\ (1,2,1) & & & (2,4,2) \\ & (2,3,1) & (3,4,1) & \end{array}$$

The probability is $\frac{6}{64}$.)

12. What is the probability that the number given by the sum of the numbers indicated by the first two pointers minus the number indicated by the third pointer is two? (Answer: The triples which work are:

$$\begin{array}{cccc} (1,2,1) & (2,1,1) & (3,1,2) & (4,1,3) \\ (1,3,2) & (2,2,2) & (3,2,3) & (4,2,4) \\ (1,4,3) & (2,3,3) & (3,3,4) & \\ & (2,4,4) & & \end{array}$$

The desired probability is $\frac{12}{64}$.)

13. What is the probability that on one spin the three pointers will *not* give a sum of 12? (Answer: We know there are 64 ordered triples in all. Only one of these gives a sum of 12. Therefore 63 of them give a sum that is *not* 12. The desired probability is $\frac{63}{64}$.)

The last question above suggests another general principle of probability. If you have selected a set of dots corresponding to an event A, there is another event B corresponding to *all the rest of the points*. If there are K points in set A, then there are 64-K points in set B. The probability of A is $K/64$ and the probability of B is $(64\text{-}K)/64$. That is, $Pr(A) + Pr(B) = 1$. Since B contains all the points not in A, the event B is the event *A does not occur*. We can abbreviate by saying that event B is the event *not-A*. Then $\Pr(A) + \Pr(\text{not-}A) = 1$. Event A and event not-A are called *complementary events*. The results we have ob-

tained here are not peculiar to a situation in which there are exactly 64 possible outcomes. Since any event either occurs or does not occur, it must be that:

(V) For any event E, $Pr(E) + Pr(\text{not-}E) = 1$

For teachers who are introducing some of the notions of set theory, we are dealing here with the concept of the *complement* of a set. If you have a *universe* of points with which you are dealing (here the 64 points in the cube) and a given set A in this universe, the set called the *complement of A* consists of all the points in the universe which are not in A. Principle V above is a statement in probability language that for any set A, the union of A and the complement of A is the entire universe.

Teachers who are using work in probability to support more difficult work in arithmetic than is exemplified in the previous 13 questions can construct questions which entail much more arithmetic such as:

14. What is the probability that on one spin the three pointers will land so that if the number on the first pointer is multiplied by itself and the result is then multiplied by the number from the second pointer and from this result is subtracted the number given by the third pointer, the result will be 14?

INDEPENDENT EVENTS

Thus far we have dealt with probabilities where two dice or three pointers were concerned. Suppose, as a change, we throw one die and one coin. If you want to avoid dice and coins, you could use one two sided checker and one pointer, or any other two convenient devices. When you throw one die and one coin, using 'H' for *heads* and 'T' for *tails*, the following ordered pairs are possible

$$\begin{array}{cccc} (H, 1) & (H, 4) & (T, 1) & (T, 4) \\ (H, 2) & (H, 5) & (T, 2) & (T, 5) \\ (H, 3) & (H, 6) & (T, 3) & (T, 6). \end{array}$$

If the coin and die are symmetric, these 12 possible outcomes are equally likely and each has probability $1/12$. As in the preceding examples you can now ask for the probabilities of many events. For example, what is the probability of *getting Heads and getting an even number*? There are three ordered pairs that work: $(H, 2)$, $(H, 4)$, $(H, 6)$, and the probability is $3/12$.

Now consider the following examples. If you throw just one coin,

$$Pr(\text{tails}) = \tfrac{1}{2}.$$

So the *multiplication rule* for finding the probability of the combination of two events does not work here either.

You have seen that for some kinds of events A and B the rule $Pr(A \text{ and } B) = Pr(A) \cdot Pr(B)$ works and for other kinds it does not. This distinction leads to an important definition.

An event A and an event B are said to be *independent events* if

$$Pr(A \text{ and } B) = Pr(A) \cdot Pr(B).$$

We can turn our definition of 'independent events' around into another important principle:

(VI) If A and B are independent events,
then $Pr(A \text{ and } B) = Pr(A) \cdot Pr(B)$.

The definition of *independent events* above does not help you directly to decide when events involving dice, pointers, or coins can be treated as independent. It tells you how to handle the probabilities provided that you know two events are independent. The 6-by-6 chart that you made for a throw of two dice assumed that events involving one of the dice are independent of events involving the other. For example the event A = *getting a 4 on the first die* and the event B = *getting a 1 on the second die* are independent events because $Pr(A \text{ and } B) = Pr(A) \cdot Pr(B)$ that is, $\frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$. Similarly the event C = *not getting a 5 on the first die* and the event D = *getting a number larger than on 2 the second die* are independent events. If you look at the 36 possible outcomes from a pair of dice, you will see that 20 of them have first die not n 5 *and* second die larger than 2. Thus you know that $Pr(C \text{ and } D) = \frac{20}{36}$. But $Pr(C) = \frac{5}{6}$ and $Pr(D) = \frac{4}{6}$. Consequently events C and D are independent events.

Deciding in advance when two events are independent resembles somewhat the problem of deciding what things in the physical world are *straight enough* that you can think of them as straight lines and apply Euclidean geometry to them. It is a problem of judgment and common sense to decide when a mathematical system can be applied profitably to a physical situation.

Physical events which do not have anything to do with each other, which do not affect or influence each other, are the kind to which we apply the mathematical idea of independent events.

If you throw two dice in the usual way, you are confident that the outcome from one die does not influence the outcome from the other die. If you glue a red die and a white together and *throw* them, you know the outcome on one die greatly influences, in fact it determines, the

illustrated in Figure 29. (In Figures 29 to 33 the white die is always on the right.) The spring is in a *neutral* position when both faces show any of the arrangements illustrated in Figure 30. If the spring *winds* a quarter turn as the dice land, the arrangements in Figure 31 show. If the spring

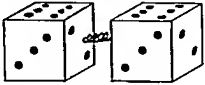


FIG. 29

FIG. 30

FIG. 31

FIG. 32

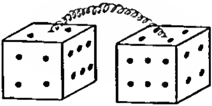


FIG. 33

unwinds a quarter of a turn as the dice land, the arrangements in Figure 32 show. Let us assume that the spring very rarely turns more than a quarter-turn in either direction and very rarely lands the way illustrated in Figure 33. If such a device were constructed and *thrown* 62 times, a chart of the results might look like the one illustrated in Figure 34. The general *diagonal* nature of the chart indicates that the separate events on the red die and on the white die are not independent. Again (from the chart alone) we say that red-die events and white-die events

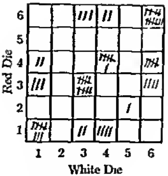


Fig. 34

are probably not independent since such a chart is possible but highly unlikely if two ordinary dice are thrown. If we throw an ordinary pair of dice 62 times and then throw it again 62 times and again and again . . . , the chart above will occur in extremely few of the trials. From our knowledge of the mechanical connection between the two dice, we are sure that events on the two dice are not independent. A measure of the dependence of occurrences such as these is called their *correlation*. When the two dice are rigidly fastened together, the correlation between them is 1; when two dice are thrown in the ordinary fashion, the correlation is 0. (For a treatment of correlation, see the references at the end of this chapter.[2,3,4,5,6,7])

PREDICTING INDEPENDENT EVENTS

Suppose you have a box the inner workings of which are unknown to you. From the outside of the box there are visible two dials and one *handle* (Fig. 35). When you push down on the handle, the two dials

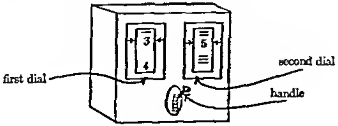


Fig. 35

spin and eventually each dial stops at a position corresponding to one of the whole numbers 1 through 6. You cannot examine the mechanism inside but must predict whether occurrences from the first dial are independent of occurrences from the second dial. All you can do is push the handle and record what happens! Suppose 30 tries give you the

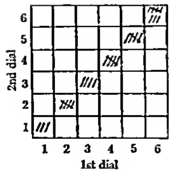


FIG. 36

results illustrated in Figure 36. You would be quite sure that the two dials were tied together somehow and you would predict with considerable safety that first dial events and second dial events were not independent. There is a slim chance that you could be wrong and that occurrences on the two dials are independent. What are your chances of being wrong? Suppose that readings from the dials *are* independent and that each number has a probability of $\frac{1}{6}$. What are the chances that one push of the pointer will land you in the *diagonal of doubles*? This is a problem just like dice. Your chance of landing in any one square of the chart is $\frac{1}{36}$ so the chance of landing in one of the six squares along this diagonal is $\frac{6}{36}$. What are the chances that in two spins each result will land in the diagonal? Under the assumption that spins are independent events, the probability is $\frac{1}{6}\cdot\frac{1}{6}$. What are the chances that each of 30 throws will land in this diagonal? $\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\ \ldots\ \frac{1}{6}$ with 30 factors, or $(\frac{1}{6})^{30}$ which is about 10^{-23}. So, when you predict that results on the two dials are not dependent you are rather safe. You are saying that either an event of probability 10^{-23} has occurred or your prediction is correct!

Usually you do not work with overwhelming probabilities such as 10^{-23} when making decisions and predictions. A much more difficult problem than the diagonal chart for the black box is to consider a chart in which the occurrences tend to fall all over the chart. To predict, then, whether the two dials are interconnected so that they influence each other requires more machinery than we have yet developed. Such problems are one type dealt with in statistics.

The black box with two dials is a simplification of many of the problems men actually deal with. For example, there are men on both sides of the argument as to whether the occurrence of sunspots is linked up to the business cycle on earth. The argument is really over whether these two kinds of events are independent.

OTHER STUDENT ACTIVITIES

Go back to the idea of six sided lettered cube which *lands* on each of its six sides with equal probability. Suppose also that the sides of this cube are labelled respectively with the letters

$$T \quad F \quad E \quad I \quad O \quad N.$$

If you toss this cube three times, you will get three letters in order, first letter, second letter, and third letter. What is the probability that, in that order, you will get the letters N, E, T, that is, that you will spell the word *net*. Assuming (justifiably) that these are independent events and knowing that the probability of each letter separately is $\frac{1}{6}$, the probability of spelling *net* in order is $\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = \frac{1}{216}$. (It would be the same problem if the three letters were obtained by tossing three cubes once.)

Now suppose that you are allowed to rearrange the three letters in order to spell *net*. The outcomes which will lead to success upon rearrangement are

$$\begin{matrix} N & E & T \\ N & T & E \\ T & N & E \\ T & E & N \\ E & N & T \\ E & T & N \end{matrix}$$

and since each one of them has probability $\frac{1}{216}$, the probability of *net* allowing rearrangement is the sum of the separate probabilities (the events are mutually exclusive) $\frac{1}{36}$. For older students problems such as this can be as involved as you choose to make them:

What is the probability, allowing rearrangement, of getting "finite" with six cubes?

The probability without allowing rearrangement is $(\frac{1}{6})^6$. A student who lists all of the ways that the cubes may fall in order so they can be rearranged to *finite* will be ready to study *permutations and combinations** to learn fast methods of finding the number of successes. One warning: in making such a list of successes, one '*i*' and the other '*i*' need to be thought of as different letters—think of a red '*i*' and a white '*i*'. Then here are two different successful tosses

$$I_r \quad F \quad I_w \quad N \quad T \quad E$$
$$I_w \quad F \quad I_r \quad N \quad T \quad E.$$

* See almost any text in college algebra.

There are $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ successful sequences so the probability of getting *finite*, allowing rearrangement, is

$$\frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6^6}$$

which is approximately .015. Students who are specialists in *word games* can find the probability that, say, five cubes tossed will give some word, either with or without allowing rearrangement (depending on how hard a problem is desired).

EQUATIONS WITH RANDOM COEFFICIENTS

One of the teaching problems connected with solving equations, as with many other topics in mathematics, is to give students adequate practice while at the same time keeping their interest. A list of 100 equations to solve can be so dull that the student works through them almost without thinking about them and therefore with small benefit. Here is one suggestion for *interesting drill* via probability.

Suppose you have an equation pattern such as

$$\square x + 2 = 14.$$

If you put a numeral, say '3', in the box, then you have an ordinary equation in one variable:

$$3x + 2 = 14.$$

Now instead of choosing a 3, think of throwing a die to find what numeral to put in the box. There are six possible equations and each one has its root. The root depends on what happens when the die is thrown. What is the probability that the root is 3? The six possible equations and their corresponding roots

$x + 2 = 14$	root: 12
$2x + 2 = 14$	root: 6
$3x + 2 = 14$	root: 4
$4x + 2 = 14$	root: 3
$5x + 2 = 14$	root: $\frac{12}{5}$
$6x + 2 = 14$	root: 2

each have probability $\frac{1}{6}$ (because the die is symmetric) and so the probability of *getting 3 as a root* is $\frac{1}{6}$.

Now consider the pattern

$$\square x + \triangle = 14.$$

Now throw two dice, one to give the numeral for the box and the other

to give the numeral for the triangle, and ask for the probability of getting 2 as a root.

In order to find this probability the student must find the roots of 36 equations! Moreover, these equations are conveniently constructed to help students develop accurate intuition for what happens to the roots of an equation as the coefficients are changed. The 36 equations have the following as roots:

$$\begin{array}{cccccc}
13 & 12 & 11 & 10 & 9 & 8 \\
7\frac{1}{2} & 7 & 6\frac{1}{2} & 6 & 5\frac{1}{2} & 5 \\
\frac{13}{3} & 4 & \frac{11}{3} & \frac{10}{3} & 3 & \frac{8}{3} \\
\frac{13}{4} & 3 & \frac{11}{4} & \frac{10}{4} & \frac{9}{4} & \boxed{2} \\
\frac{13}{5} & \frac{12}{5} & \frac{11}{5} & \boxed{2} & \frac{9}{5} & \frac{8}{5} \\
\frac{13}{6} & \boxed{2} & \frac{11}{6} & \frac{10}{6} & \frac{9}{6} & \frac{8}{6}
\end{array}$$

Each equation has probability $1/36$ and since 2 occurs as a root three times, the probability of getting 2 is $3/36$ or $1/12$. Other questions based on this same equation with two random coefficients might be:

What is the probability of getting more than 2? ($27/36$ or $3/4$).

What is the probability of an even whole number as a root? ($8/36$ or $2/9$).

What is the probability of getting 1 as a root? (0).

Although equation solving is usually a high school activity it can (and should) be taught in elementary school as soon as the child knows the corresponding arithmetic.[8] Then work with random coefficients can be developed as understanding of probability develops.

After a high school student has done some work with simple equations in one variable, he can consider systems of equations with random coefficients. For example, consider the pattern

$$\begin{aligned} \square x + \triangle y &= 100 \\ 4 \cdot \bigcirc x + \Diamond y &= -50. \end{aligned}$$

Assume that the random coefficients have probabilities as follows (students have outgrown the dice):

$\square$ takes on the value	With probability
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

(Notice the chances for $\square$ are just the same as for dice.)

○ takes on the value	With probability
5	$\frac{1}{2}$
0	$\frac{1}{2}$

△ takes on the value	With probability
0	$\frac{1}{3}$
$\frac{1}{2}$	$\frac{1}{3}$
1	$\frac{1}{3}$

◇ takes on the value	With probability
3	1

In order to answer questions concerning the probability of ordered pairs (x, y) as roots for this system of equations, a student must solve 36 separate systems of equations. Typical questions to be answered might be:

What is the probability that the sum of the numbers in a pair is positive?

What is the probability that $x > y$?

What is the probability that if a root (x, y) is interpreted as giving the coordinates of a point, that point will fall outside of the unit square (Fig. 37)?

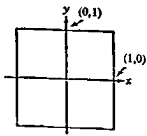


FIG. 37

Only a few of the elementary ideas of probability have been touched on here. It would take another chapter just to list all the important topics not covered here. However, a student who has worked with some of the suggestions given here will have a conceptual advantage when he does more work with probability later, either in formal courses or in self-study. More important, preliminary work with probability as described here may add a bit more zest to the student's study of mathematics in the first twelve grades.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials of this chapter.

NOTES

1. Cohen, John. "Subjective Probability." *The Scientific American* 197: 128–38; November 1957.
2. Cramer, H. *The Elements of Probability Theory and its Applications*. New York: John Wiley & Sons, Inc., 1955.
3. Feller, W. *An Introduction to Probability Theory and its Application*. New York: John Wiley & Sons, Inc., 1950, Vol. 1.
4. Commission on Mathematics. *Introductory Probability and Statistical Inference for Secondary School*. An Experimental Course. Preliminary edition. New York: the Commission, College Entrance Examination Board, 1957.
5. Kemeny, J.; Snell, J.; and Thompson, G. *Introduction to Finite Mathematics*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1957.
6. Neyman, J. *First Course in Probability and Statistics*. New York: Henry Holt & Company, Inc., 1950.
7. Robbins, Herbert. "The Theory of Probability." *Insights Into Modern Mathematics*. Twenty-Third Yearbook. Washington, D.C.: The National Council of Teachers of Mathematics, 1957.
8. UICSM Staff. "Arithmetic with Frames." *The Arithmetic Teacher* 4:119–24; April 1957.

7

Statistics

RICHARD S. PIETERS AND JOHN J. KINSELLA

"It is truth very certain that, when it is not in our power to determine what is true, we ought to follow what is most probable."—René Descartes[1]

We would probably fail if we searched courses of study and textbooks of the first six grades in the hope of finding any specific reference to statistics. Even in junior and senior high schools little if anything is done in this field at this time. Since statistics has never been an essential part of the traditional course in mathematics in either the general curriculum or the college preparatory curriculum, why should a chapter on it be included in this book? After all, even adults have difficulty comprehending data expressed in statistical form. Why should we expect youngsters of school age to possess the readiness for such material?

Before we can give even partial answers to these questions about the place of statistics in schools and the ability of students to comprehend it, we should say what modern statistics really is. Essentially it is a method for solving problems which involve making decisions in the face of uncertainties due to incomplete information. This involves, of course, deciding what data are pertinent to the problem, collecting as many data as are feasible, organizing and presenting them, interpreting them, and finally, using them to make the decision which is most likely to be correct.

However, this analysis might be applied to almost any kind of problem. The special feature of statistical problem solving is the use of the *theory of probability* in choosing the method of *sampling* the data and in drawing conclusions about the *population measures* from the *sampling statistics*. This special characteristic of modern statistics has been clearly stated by many writers. "In the theory of probability we deduce the probable composition of a *sample* from the composition of the original *population*. But statistics builds on the theory of probability, making it possible for us to reverse the reasoning—that is, we infer the

composition of the original *population* from the composition of a properly chosen *sample*."[2] When a national magazine asked a *sample* of citizens a series of questions about their reactions to the success of the United States in getting its first artificial satellite into outer space, it was trying to find out what the reactions of all of us (the *population*) were to the important event.

THE IMPORTANCE OF STATISTICS

Our first question was whether statistics should be given more attention in our schools. The answer is more likely to be "*yes*" if it is apparent that a knowledge of statistics is rapidly becoming more and more essential to both the general education of all citizens and to the vocational preparation of an increasing number of future specialists. From a man who has made outstanding scholarly contributions to the theory of statistics, worked assiduously to make its applications known to men of practical affairs, and, as a teacher, thought seriously about the problems of statistical education, we have this judgment: "I do not think it is an overstatement to say that probability and statistical concepts are involved in the problems and thinking of modern scientific and technological society as much as or more than any other body of mathematical ideas. At one end of the spectrum we find the average citizen confronted with the intelligence scores of his children, insurance problems, advertising and sales claims, public opinion polls, etc. He should be introduced to at least the rudiments of probability and statistics at the high school level. At the other end of the spectrum there is the scientist in almost every field who designs his experiments and analyzes and interprets the results by probability and statistical methods."[3] Modern statistics is used in the design of agricultural experiments to help us find ways of obtaining more food and better food at less cost.[4] Medical research uses statistics to find out whether certain treatments, like polio shots, reduce or prevent the incidence of certain diseases.[5] Business and industry use statistical methods to check the quality of their products and to decide on the most economical ways to organize their operations and processes.[6] Engineers and production managers of the future must know the probability and statistics basic to *operations research* of which *quality control* is an example.[6] The future psychologist must have a background in probability and statistics both for research design and for understanding learning models based on probability.[7] The biologist who wants to be competent in the field of genetics had better be acquainted with *Markov chains*, which in turn leans on probability concepts.[7] The future development of the *theory of games*, also dependent on probability ideas, is

likely to yield more applications in devising strategies in military affairs and in business competition.[7] Statistical mechanics is already a standard course in the preparation of engineers and physicists. The astronomer is applying the theory of probability to the statistical study of the distribution of star galaxies.[8] The *Monte Carlo* method, an offspring of probability, has been successfully used in the study of the neutron.[9] Such diverse subjects as heat and information theory seem to be governed by similar laws of probability. It is well known that the insurance companies depend on laws of probability in dealing with the duration of human life.

The question about the ability of precollege students to understand certain statistical ideas, not given much attention at present, can only be answered by classroom experimentation. In our opinion many of these concepts are simpler than some mathematical topics already in the curriculum. The strangeness of some of the new ideas can be removed by the abundant use of familiar illustrations. It is a primary purpose of this chapter to make these educational hypotheses of ours more plausible, and to encourage classroom experimentation with the statistical ideas we present.

HOW MUCH STATISTICS IS NOW TAUGHT

If we ask whether the development of problem-solving ability is one of the principal concerns of the elementary school, we would almost certainly be answered in the affirmative. If we inquire whether the collection of data is part of the process of solving problems, we would probably be granted another "yes." If we follow up our two successes by innocently asking whether any of the data gathered are in numerical form, it is quite likely that we would receive an "of course" reply. Since the *collection* of numerical data to solve problems is part of so-called *descriptive statistics*, we would almost have to conclude that statistics in a crude form had already infiltrated the elementary school.

It seems to us, however, that more attention could profitably be devoted in the first six grades to the *organization* of data into tables and graphs. Attendance records, heights of children, and so on might be appropriate data for collection and organization. The median as a mid-score and the mode could be introduced quite early in the grades to obtain single, representative scores for a group of scores. After division had been learned, the average known as the mean (arithmetic mean) could be applied to data to obtain another kind of measure of central tendency. Even the effect of large and small scores on the mean could be studied, as well as the stability of the median and mode under similar treatment.

As is pointed out in Chapter 6, (page 247) the relative frequency concept of probability as the ratio of the number of *cases favorable to the occurrence of an event,* to the *total number of possible happenings,* could be used after fractions had been presented. Numerous familiar illustrations, such as the probability of a girl's name being drawn from a class list by chance, could be used to develop the concept. Questions of the form "What are the chances that . . .?" applied to a variety of such examples would precede the introduction of the word *probability.* Weather information about the number of stormy, cloudy, and fair days in a given month could also be used in illustrating the concept.

If we next turn our attention to the junior high school, we note that the ability to interpret data in tables and graphs is frequently a serious objective of instruction. However, not much attention is given to *frequency graphs,* such as the *histogram,* the *frequency polygon,* the *cumulative frequency graph,* and the *graph of cumulative per cents.* These would not be difficult to teach. These are the kinds of graphs which are concrete and basic to developing more sophisticated statistical notions. With them is initiated the concept of *distribution* and the idea that the area under certain curves may be used to represent either the total frequency or a per cent of the total frequency.

Unfortunately, we would find that these junior high school students are rarely given experiences in starting with a question or problem, deciding which data need to be collected, planning how to organize the data, and actually collecting them. All that we would usually see would be *canned* data in tabular or graphic form. The only tasks left for the students are to read and interpret them. We think that it would be possible and desirable to give these students experience with the whole problem-solving process instead of just its terminal operation.

It seems to us that histograms should be introduced not later than the 7th or 8th grades of the junior high school. The histogram is merely a vertical bar graph in which the areas of the bars of constant width are proportional to the frequencies. Along the horizontal axis are placed the measurements or categories of the variable whose frequencies are being studied. The frequency of measurements based on the body, like the span of a student's hand or the length of his stride, might be used to study the distribution of different sizes in the class, or, in the case of categories, the frequency of different eye colors in the group might be surveyed. Questions about the most frequent or least frequent value of the variable could be framed. The probability notion (Chapter 6, page 265) could be given attention by asking the students what the chances are that the name of a student with blue eyes, for example, might be drawn from a well-shuffled pack of class cards. The probability idea could be extended

by asking what the chances are that the name of a blue-eyed *or* a brown-eyed student be drawn. Questions of this type lead to the summing of probabilities to answer "or" questions, and, under proper guidance, to the obvious fact that the sum of the probabilities of an entire set of mutually exclusive events equals 1. Incidentally, the students can be provided with further experience in choosing appropriate scales to represent data, and, in the case of the histogram, learn the importance of always specifying the location of zero on the vertical axis. Finally, the tables which provide the raw data for the graphs can be extended to show in decimal form what proportion each frequency is of the total frequency (see Table 1, p. 297). A partial check on these computations is that their sum should equal 1. There is also an opportunity here to get the students used to seeing probabilities expressed in decimal form.

From the histogram it is easy to obtain the *frequency polygon* for the same data by drawing a broken line graph through the midpoints of the top bases of the vertical bars, and connecting the endpoints to points on the horizontal axis one-half unit to the left and right of the limits of the histogram. (Fig. 1.) It is not difficult to get students to see that the total area of the bars equals the area under the frequency polygon. Only some intuitive notions of congruent right triangles are needed.

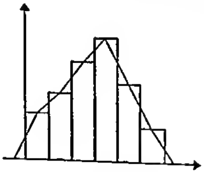


FIG. 1

It is a common practice in probability and statistics to assign the area under a graph the value 1 and use this geometric picture to represent probability measure, just as it is common in the calculus to picture the derivative as the tangent of the angle of inclination that a line tangent to a curve makes with the x axis. By selecting various areas under the polygon the teacher can give the students experience in interpreting this area-probability analogue.

By choosing data, like heights of students, the graphs of cumulative frequency and cumulative per cent can be taught. Questions about how

many and what per cent of the students are taller or shorter than a given height can be posed and answered from the graph. Furthermore, the use of cumulative per cents gives a natural introduction to the percentiles, deciles, and quartiles, including the median.

This also gives an excellent opportunity to point out that the average height or weight of a group of normal students is not necessarily the height or weight of a normal student. In fact, variability from the average is much more likely to occur than identity with it. Not only should we be acquainting children with the uses of statistics but also we should be pointing out to them some of the misuses and misconceptions of statistics which are all too common.

How To Lie with Statistics is the title of a 1954 book by Darrell Huff[10] which brings out very cleverly some of these misuses. In their *Statistics, a New Approach*, Wallis and Roberts devote a whole chapter to misuses of statistics.[11] They classify some of these misuses as due to (1) shifting definitions, (2) inaccurate measurement or classification of cases, (3) methods of selecting cases, (4) inappropriate comparisons, (5) shifting composition of groups, (6) misinterpretation of association or correlation, (7) disregard of dispersion, (8) technical errors, (9) misleading statements, and (10) misleading charts. Beginning not later than the 8th grade teachers of mathematics should seek to make their students more critical of some of these distortions of the truth. One of the best procedures is for the teacher and the class to clip from newspapers and magazines material whose conclusions or methods of arriving at conclusions seem doubtful. Not all of these will be useable, for frequently the material being discussed requires special knowledge or a degree of sophistication not yet attained by the students.

Under *misleading charts* Wallis and Roberts show how a broken line graph might be used to exaggerate an increase in the cost of living. This was done by choosing a few years during which the increase occurred and ignoring earlier and later years, and by magnifying the vertical scale. In another example the purchase and redemption of U. S. bonds was compared by using a scale for the redemption data three times as large as that for the sales. Other misleading devices are to omit the zero point on the horizontal or vertical scales, and to use the same intervals on the horizontal scale for units of different size.

In the elementary algebra course of the ninth grade we find that in some schools a unit of descriptive statistics has been inserted, probably just before the graphing of linear equations. This unit usually involves tables, histograms, frequency polygons, the arithmetic mean, the median, the mode, and the range. But this material is often an island unit

in that its concepts do not appear later in any of the other units of instruction.

Ninth grade students may be asked to write a formula for the mean. There may be *story problems* about finding the score or grade needed to obtain a desirable scholastic average, but the rapid method of finding a mean by taking deviations from an arbitrary score is rarely mentioned.

Let us stop a moment to illustrate this useful method by an example. Suppose that the amount of change in the pockets of five boys is 23¢, 36¢, 19¢, 31¢, and 21¢, respectively. The problem is to find the mean of the five amounts. Suppose we guess the mean to be 25¢. Subtracting 25 from each of the five numbers we obtain -2, $+11$, -6, $+6$ and -4. The sum of these five deviations is $+5$. Their mean is $+5/5 = +1$. Adding this mean of the deviations to the guessed mean, we get $(+1) + 25 = 26$. This is the mean of the five amounts, for

$$\frac{(23 + 36 + 19 + 31 + 21)}{5} = \frac{130}{5} = 26.$$

Getting class averages this way gives useful drill in simple operations with signed numbers.

It seems that the mean equals the guessed mean added to the mean of the deviations. If we want to prove this for better students it may be done as follows: Symbolizing the mean by $\bar{x}$, the guessed mean by g, and the mean of the deviations by $\bar{d}$, we want to prove that $\bar{x} = g + \bar{d}$. To show this for any five scores, represent them by x_1, x_2, x_3, x_4, and x_5. Then

$$\bar{d} = \frac{(x_1 - g) + (x_2 - g) + (x_3 - g) + (x_4 - g) + (x_5 - g)}{5}$$

$$= \frac{(x_1 + x_2 + x_3 + x_4 + x_5) - 5g}{5}$$ by the associative and commutative laws

$$= \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} - \frac{5g}{5}$$ by the distributive law for division

$= \bar{x} - g$ by substitution of equals.

Hence, $\bar{x} = g + \bar{d}$ by the addition axiom for equations. The same argument can be used with any number of scores. A further important observation is that if the real mean is equal to the guessed mean, $\bar{x} = g$, then $\bar{d} = \bar{x} - g = 0$ by substitution of equals. This means, writing d_1 for $x_1 - g$ etc., that $\frac{(d_1 + d_2 + d_3 + d_4 + d_5)}{5} = 0$ by substitution of

equals, and that $d_1 + d_2 + d_3 + d_4 + d_5 = 0$ by the multiplication axiom for equations. We now have proved that the sum of the deviations from the mean of five scores is zero. Again, this result can be proved for any number of scores by the same kind of argument.

FURTHER CONCEPTS OF STATISTICS IN THE SCHOOLS

Symbolism is an important concept in the study of statistics as in many fields of mathematics. But even among college students inability to interpret symbols is a serious block to their progress in mathematics. Some of the casualties might be prevented if experience with a variety of symbols was given during the secondary school years. The symbols used to represent variables and constants in elementary algebra are pretty well confined to lower case letters and capitals of the Roman type. There is practically no use of subscripts, primes, and Greek letters which are used so frequently in statistical formulas.

Approximation is also an important idea in statistics. Not only is this reflected in the way probability deals with uncertainty but also in the fact that the data of statistics are frequently *measurements*. We know that all measurements are approximate. If the heights of ninth grade boys are being examined, a score of 65 inches could not be represented by a point. Instead, a line segment extending from 64.5 to 65.5 would be necessary. If a mean height is to be determined from such data, the principles of computing with numbers which represent approximations must be known. Evaluation of mensurational formulas, involving the addition, subtraction, multiplication, division, and square root of such numbers, and other geometrical and trigonometrical problems in which attention must be given to the concept of approximation should be emphasized.

Discreteness and *continuity* are essential ideas in statistics. Dots, and unbroken lines and curves are the respective graphic pictures of these. In the past, we have in ninth grade algebra either restricted ourselves to the continuous kind of variation, or have blithely assumed continuity for variables which are discrete. We should use more relations which require dot diagrams for their study. We should have the students clearly define the domains and ranges of variables by means of sets.

In the tenth grade we concentrate on *deduction* as the method of mathematical proof. The conclusions are certain if correct logic is applied to the accepted definitions, postulates, undefined terms, and previously proved theorems. *Induction* is used sparingly to arrive at hypotheses to be tested by deduction, and the danger of basing conclusions on it is emphasized.

Among the ten misuses of statistics listed by Wallis and Roberts were *shifting definitions*, *methods of selecting cases*, and *misleading statements*. The tenth year with its emphasis on logical thinking in nonmathematical as well as mathematical situations is a good time to spike these misuses of statistics. For example, when unfriendly nations criticize the amount of unemployment in the United States, it is to their advantage to shift the definition of number of unemployed to mean all of those workers who do not work full time.

In the eleventh or twelfth grade a unit on *permutations and combinations* is followed by one on *probability*. The latter is defined in terms of *relative frequency*. Problems of the *a priori type*, in which the probability of an event is *assumed* in advance of an experiment, engage the attention of the students. Examples such as coin tossing, dice throwing, card drawing, and lifting balls from an urn are used. Too many of these problems involve the equal probability of events. Life insurance tables are used to illustrate the necessity of *a posteriori* probability.

The modern concept of probability as a measure of a subset of the set of all possible events of a certain kind does not appear. The postulational approach to probability is avoided. Independent events and mutually exclusive events receive considerable attention but conditional probability is usually concealed behind *dependent* events. The binomial expansion is seldom related to mutually exclusive events. Due in large measure to the lack of systematic treatment, the students find the textbook problems very frustrating.

FUNDAMENTAL IDEAS AND TERMS

We are all aware of the increase in the birth rate of the United States since 1942. In certain sections of our country this population growth was much more rapid than in others. As this phenomenon continued, the future need for more classrooms and school buildings became evident. In many communities the local school boards reported the data to citizens through the medium of tables and graphs.

The collection, organization, presentation, and interpretation of these numerical data are *descriptive* statistics. The term statistics has also been applied to the data themselves.

The process of counting was the essential one in collecting birth data. However, we can see that in planning for the location of a school and for the cost of transportation of students to it, the distances from the homes to the prospective school had to be taken into account. Hence, *measurement* scores as well as frequencies were involved in the statistical treatment

If we were concerned about this problem in our own community we would probably try to collect information about *all* births over a period of years. In other words, we would collect *population* data; we would hardly be satisfied with a partial census. On the other hand, if we wanted to estimate the *national* need for new schools, we almost surely would have to deal with a part of the population. In this case we would select a *sample*. From such limited data we would hope to predict with an appropriate degree of accuracy the national need for additional classrooms.

This is a problem in *induction*. From the information given by the sample we would want to obtain certain tentative generalizations about the population of which the sample is a part. A major problem is how to *select* such a sample so that our estimate of the population characteristics possesses adequate accuracy.

In selecting the sample we would certainly want to eliminate *bias*. In other words, we would want to be sure that we did not draw our sample from a very special part of a population, and then attempt to draw conclusions about a whole population. For example, in drawing a sample on a national basis we would hardly limit our sampling to states east of the Mississippi, or to cities, or to the wealthiest states, or the coastal states.

We would also be concerned about the *size* of our sample. What per cent of the population should constitute our sample? If we chose a very small per cent, we might worry about the reliability of our prediction, for small samples are likely to vary much more than large ones. If we selected a very large sample, the cost of collecting the data might exhaust our financial resources.

Even if we were satisfied with the size of our sample, we would still have the problem of determining the best way to use the information to arrive at sound estimates of the population's characteristics. Unless we had used *randomness* somewhere in the process of selecting the sample, we would be blocked. For example, we might divide each state into geographical areas and in some random manner select communities from within each of these areas. If we had used the random process in the appropriate way, we would have satisfied one condition for using *probability theory* to help us make our estimates.

A second condition is that we must know the *distribution* of certain measures calculated from the sample data. For example, we would probably want to know how the birth rate varied from one sample of the population to another. In some samples this rate would be relatively small, and in others fairly large. The essential information needed is how frequently birth rates of different sizes would occur if all possible samples

had been drawn from the population. If the relationship between the size of the birth rates and their frequency is or at least approximates a known mathematical function, like the bell-shaped normal curve, the problem of making the population estimates can be readily solved. When a known distribution does not appear, other more complicated statistical methods may be found applicable.[12]

THE RELATION BETWEEN PROBABILITY AND STATISTICS

Under conditions like these probability teams with descriptive statistics when inferences are to be made from a sample to the population of which it is a part. Whereas the statistics of 30 to 40 years ago used to be limited pretty much to descriptive statistics, modern statistics is a combination of descriptive and inferential statistics in which probability is an integral part. Statistics and probability can no longer be separated.

We have been talking about the application of probability to statistical data. Certainly probability as a branch of mathematics has an existence independent of its applications.[13] This aspect of probability is treated in the preceding chapter.

Probability and statistics, as pure and applied mathematics respectively, exhibit all of the themes or all-pervading concepts discussed in other chapters of this book. Both real and complex numbers, and of course the operations on them, are involved in the theory and applications of probability and statistics. We have already illustrated the ways measurement and approximation permeate the union of the two subjects. Induction and deduction are both involved. Relation and function are "on stage" whether we are using tables and graphs, or considering functions like the normal curve, the *t*-curve, chi square, or regression lines. Symbolism, the bane of novitiates in the field, involves sigmas, large and small letters, Arabic and Greek letters, subscripts, and integral signs, among others.

TWO ILLUSTRATIVE PROBLEMS

Many children are interested in baseball. What does a man's batting average mean? We may say that because John's average is .375 and Sam's .253, John is a better batter than Sam. Does this mean that if both John and Sam bat four times in a certain game John is sure to get one and a half hits and Sam one? Obviously not. Maybe Sam will be lucky in that game and get three hits while John will have a bad day and go hitless.

Suppose, as a second illustration, someone is interested in knowing how many left-handed students there are in your class of 40 students.

How would you find out, and how would you convey the information to the inquirer? The answer to this question is easy and obvious, but if the question were asked in a school of 1000 students, how would the problem be tackled? And what if it were asked in a school system like New York City's with nearly a million students?

Now the question might not only concern the total number of left-handed students in the system, but also how many could be expected on the average in each class of 40 students. You may ask why anyone should be interested in such problems. Well, in the first case, it might be a manufacturer of sporting goods trying to decide how many baseball gloves to make for left-handed boys or how many sets of left-handed golf clubs are needed; or in the second case, it might be an architect trying to decide a question of equipping the classrooms in a new school with the proper number of left-handed writing chairs. Then the fact that one given class happened to have three left-handed students might be interesting but would hardly be helpful to the architect, particularly if in other classes in the same school he found that five had no left-handed pupils, three had one each, and one had twelve. The study of a problem like this should be based on statistical considerations. No amount of algebra or geometry, arithmetic or trigonometry by itself is going to help.

Suppose the architect determines that in order to build the school to best advantage he needs to solve this problem. How does he go about it? Are the methods he uses those which youngsters can understand, appreciate, and perhaps learn to apply to similar situations?

STATISTICS AND UNCERTAINTIES

It is easy to see that there are uncertainties involved here. Suppose in the school of 1000, 50 left-handed children are found. Does this mean that in each room of 40 seats exactly two left-handed chairs will be needed? Of course not! Maybe one room will need none one year and ten the next. Is it possible that one room might need forty left-handed chairs some year? It certainly is possible, but is it probable? Far from it. After some thought, the suggestion is made that it would be wise to put in 30 right-handed chairs, 2 left-handed ones, and 8 rather inconvenient ones which can be changed from right- to left-handed, and back as needed.

Now, what are the chances that things will work out? Maybe, it would be better to have 35 right-handed, no left-handed, and 5 interchangeable chairs. But all of this seems so uncertain and so "up in the air." Is there any way we can reach some decision which is not based on pure guessing? The answer is "yes" and the methods are those of statistics.

Even though some well-formulated problems involve so much un-

certainty that there is no *unique* answer to them, *decisions* must be made every day as best we can, orders placed, and money spent. Thus, it is of the utmost importance that youngsters learn as early as possible that mathematics can be used to study problems involving *uncertainties*. Particularly, they should see that when one answer is chosen from several possible ones by statistical reasoning there is a *reasonable assurance* that it is *better* than the others.

All of this has been said many times and in many places. One statement that summarizes our discussion especially well is the following: "The notions of probability, correlation, and sampling are among the fundamentals of modern social measurement. . . . Moreover, there is in all statistics a salutary concern for the uncertain and the incomplete—for the gray that is real more than for the black and white that is abstraction. It is well for the student to learn both that mathematics has uncertainty, and that uncertainty can be mathematically treated. This knowledge is important in many fields; teachers of science and teachers of history alike have their troubles with students who are persuaded that all reasoning is geometrical and all evidence conclusive."[14]

CHARACTERISTICS OF STATISTICAL ANALYSIS

What then are the characteristics of the statement and solution of a problem by statistical methods? It seems to us that most problems subjected to statistical analysis involve decision making in the face of uncertainty, with the concomitant problem of what the chances are that the decision is right. That kind of problem is faced by the architect in choosing the kinds of seats for classrooms in a new school. The same sort of problem confronts the purchasing agent for a flour company when he has to make up his mind whether or not to buy a certain carload of wheat. The manager of a baseball team in the World Series, trying to decide whether to use a winning pitcher after only two days rest, is in a similar "forked road" situation. In the same "fix" is the buyer of mens' shirts in a large department store. What combinations of neck sizes and sleeve lengths should be stocked, and in what proportion? In the same quandary is the superintendent of a light bulb factory. What percentage of defective light bulbs can he allow before ordering the machine shut down for repairs, and how does he find out the percentage of defective bulbs it is making anyway? Even the schoolboy who claims he can tell bottled coke from tap coke will not be acknowledged as competent in that skill unless he makes correct choices a certain per cent of the time. How many times must he be right to demolish the charge that he is just lucky?

These are all problems which call for a decision. Furthermore, all of

them have a certain practicality in modern life. They are problems which are susceptible to attack by statistical methods, except that of the baseball manager, and yet what student of elementary secondary school mathematics would have any idea of how to attack them?

We shall now attempt to indicate how certain statistical ideas and processes, basic to the solution of such problems, might be given more attention in the schools.

OUTLINE OF A STATISTICAL ANALYSIS OF A PROBLEM

The first and most important idea is the one we have been emphasizing continually till now, namely, that problems of decision in the face of uncertainty are really susceptible to mathematical study. We state again the idea which may be unfamiliar to many, i.e., mathematics is not exclusively a matter of techniques of numerical calculation, algebraic manipulation, geometrical proofs and constructions, or even of analysis and deductive thinking. It may also serve as a basis for inductive inferences from partial data and incomplete information.

Too often we run across the idea that mathematics must be exact, idealistic, and certain. That a mathematical model of a physical situation may not produce conclusions that are exact and certain should receive more attention.

After we have agreed that a mathematical approach to the solution of a problem is possible, the first step is to formulate the problem carefully. It is frequently true that a careful statement of the problem yields ideas about the best method of attempting a solution, and one of the first would probably be that the situation is too large or too complex to handle *in toto*. We find therefore that we have to draw a sample from the population we are studying and we must decide how large a sample is needed and how best to draw it.

The next step would be the gathering of the relevant data from the sample, recording and tabulating them, and organizing them in such forms as graphs and tables for the purpose of presentation.

Following this we would have to interpret the data in the light of the formulated problem.

Finally, we would want to infer from the sample, with a desirable degree of probability, conclusions about the population relevant to the solution of the problem.

How many of the techniques and ideas involved in these procedures are in our present curriculum? Many of them are now there and all that is necessary is merely to point them out and emphasize that these already familiar ideas and techniques are useful in statistical problems as

well as in others. There are also some relatively simple new techniques which we hope may become part of the regular work in elementary and secondary mathematics. We recognize, of course, that many of the ideas and mathematical methods of statistics are too complex and too subtle for any but specialists in the field, and these we shall either not mention at all or merely note briefly in passing.

FORMULATING THE PROBLEM

Often the original statement of a problem is quite vague. The first requisite in an attempt at a solution should be an effort to state the problem carefully, to define the key words accurately, to recognize the assumptions being made, to disregard irrelevant facts, and so on. For example, in the problems about the architect and the left-handed chairs, the word left-handed seems to be the key word. What is meant by *left-handed*? Well, in this problem we obviously do not care if an individual bats left-handed at a baseball game or eats left-handed at dinner, but only whether, for one reason or another, he writes with his left hand. Even this stipulation is not enough, for there are left-handed individuals who write in such a position that they prefer right-handed chairs. Hence, in this instance, *left-handed* will be taken to mean *writes left-handed and prefers a left-handed chair.* With the aid of this definition we are in a better position to select the data for answering the question, "How many right-handed, left-handed, and interchangeable chairs are needed in each classroom planned for 40 pupils?"

A much more complicated situation confronts the superintendent of the light bulb factory. He really has several questions to answer before he decides whether to shut down the machine for repairs or to leave it running. Some might be: "What is a defective light bulb?", "How do I know how many defective ones I'm making a day?", "How high can I let the defective proportion rise before I act?", "Do I make this decision on the basis of profit and loss or on the reputation of our factory for producing only the best light bulbs in the industry?", and incidentally, "What does *best* mean in this situation?".

Of course, we can overemphasize the semantic aspects. However, the danger of being too slipshod in formulation is the greater one. Suppose the sales manager says that all bulbs to be satisfactory must have a guaranteed continuous life of at least 1000 hours. Now we have a definition of a defective bulb, namely, one that does not yield 1000 hours of continuous service. But how is the superintendent to know whether defective bulbs are being made? Most high school students will recognize that it would be foolish to burn each bulb for 1000 hours to find out whether it was defective. By the end of that time the useful life of the

bulb would be nearly exhausted. One result would be that the consumer would not get his money's worth.

The careful formulation and discussion of this problem by a class of students would almost certainly lead to the idea of testing a single bulb now and then from a given run. That is, *we would select a sample, test the sample, observe the results of the test, and infer from the observation a conclusion* about the quality of the bulbs being produced. Finally, *from this conclusion we would make our decision* to stop production for repairs or to continue production, even though we know the machine is making a certain number of defective bulbs which will have to be replaced at the consumer level. This is an example of *industrial quality control* which is playing an increasingly important role in many manufacturing processes.

While the importance of carefully defining words like *defective* is usually stressed during the teaching of geometry in the 10th or 11th years, we should remember that the teacher of mathematics can make a contribution to the general education of students by using illustrations and problems outside of the field of geometry to indicate the equal importance of careful definition in many other fields of mathematics as well as in a variety of life situations.

THE IMPORTANCE OF SAMPLES

Usually a problem which can be attacked by statistical methods will involve properties of a certain set of objects. It may be that the set is so numerous that it is impractical to examine every one or it may be that the property is of such a nature that its determination destroys either the object or the property. In either case we are forced to consider only a *sample* of the population we are interested in. This concept of a sample is one of the most important in statistics. We shall, therefore, spend some time discussing some of the salient properties we want in our samples and the ways samples can be selected so as to insure these properties.

The size of the sample must be sufficiently large. When samples are drawn by a random process, both experience and theory reveal that the smaller the sample, the greater is the variation from sample to sample in whatever measures are taken. This means, for example, that the error in the prediction of the mean height of all American men from the mean height of a small sample of men is much more likely to be serious than the corresponding error associated with a larger sample. We have to decide, *in advance of sampling*, the size of error we are willing to tolerate. Then we have some basis for deciding the size of the sample we should take.

Usually it is important that the method of selecting the sample should

contain in it some kind of *randomness*. The reason for this is that randomness must be present if we are going to be able to use probability theory to infer characteristics of the population from the sample.

This is the second time that the idea of *randomness* has appeared in this chapter. Perhaps it would be well to point out what is meant by randomness when referring to the selection of a sample. Intuitively it means that there has been no special selecting going on, that no favorites are being played. More technically, it means that every element in the population has the same chance as every other element to be selected as a member of the sample being chosen. Later we shall illustrate a few of the many methods of incorporating randomness into the sampling process.

SAMPLES IN THE CLASSROOM

In the next few paragraphs we shall offer some examples of sampling which might be appropriate at different school levels. Consider a class of 29 children about whom we want some information. We will try to get it by considering samples drawn in the following manner. Assign to each child a number from 1 to 29. In one box place three disks numbered 0, 1, and 2, in another box place ten disks numbered 0 to 9. By drawing a disk from each box any number from 0 to 29 may appear. After each drawing replace the disks, shake up each box and draw again. By drawing five numbers, ignoring the 00 and any number which has been drawn before, we get a sample of the class of size five. Similarly samples of other sizes could be drawn.

How might such a device be used? What kinds of questions might be asked? The key idea would be to see to what extent certain information about all the children could be obtained by taking only a sample of the class. For instance, what is the average time it takes to get to school? How close would we get to the truth if we took samples of 5? 10? 15? 20? The answers here would show what happened as sample size increased. How many hours of sleep does the class get? How long does the homework take? These two questions provide an opportunity to point out the importance of getting accurate data. What is the average height of the class? Would our answers be just as good if we took the measurements of only the first five children who entered the classroom? If we took only one row? Would such sampling give each possible sample an equal chance of appearing? In a like vein would it make any difference if our sample contained only boys or girls?

In the upper grades of the elementary school the mean of a collection of numbers is commonly obtained. What we feel is needed here is only

an extension of this experience to answer questions like those above. By such procedures it should be possible to develop inductively some understanding of the concept of a random sample, the importance of lessening the bias in samples, and the effect of the size of the sample on the accuracy of predicting the characteristics of the total group. As a less important by-product, there would be considerable practice in adding and dividing numbers. There certainly would be no difficulty in providing enough practice. For example, there are over 140,000 different samples of 5 students that can be drawn from a class of 30. This is merely the number of combinations of 5 things that can be selected from 30 things; each sample selected will differ in one or more of its members from every other sample.

At whatever level the study of functional graphs begins it could be emphasized that the points which are plotted are only a sample set of the set of all points which lie on the graph. In a linear graph a sample of size two is sufficient to determine the whole set. Is this also true for graphs of quadratic functions? Obviously not. The set of sample points must not only be greater than two in number but must be carefully selected if the sample is to be significant for the whole set. Discussion could follow as to the size of samples and how their selection should be made.

In geometry the study of locus problems leads very naturally to questions of sample points. How does one usually determine the locus of the center of a circle of radius r passing through a given point P? Certainly, one very effective way to begin the study of the problem is to take a sample of half a dozen cases out of the infinitely many possible ones to see what common property they possess. This is not a statistical problem but it will provide an alert teacher with an opportunity to discuss the notion of a sample.

A basketball coach determining the makeup of his team by watching his boys practice is using the idea of sample. The coach assumes that the reactions of his boys in practice are a good sample of their reactions in the game. Sometimes he is right but not always. In order to improve his sample he makes practice sessions as close to the real game situation as possible. A baseball player's batting average for the first month of the season is determined from this sample of his playing for the season. Whether the whole season will be consistent with the sample is always the question the coach faces. This is also the question the statistician must always consider, i.e., "Is this sample a true sample of the population I want to study?"

The trouble with these samples is that they are not really representa-

tive of the population the coach is interested in. Out of all possible reactions of a boy on a basketball floor the coach is interested in those particular reactions which occur in a game. But the sample he considers is made up of the reactions he observes in practice which is not in fact a sample of game reactions, but only of the total reactions and not even a very good sample of that because game reactions have been excluded.

BIAS IN SAMPLES

In order that the theory of probability can be applied to statistical data it is necessary that every possible sample that can be drawn has an equal chance of being selected. When the method of selecting samples does not satisfy this condition, we say that the sampling process is biased. Sometimes it is obvious that a given sample is biased. One example is the one in the last paragraph. As another, we note that it is usually true that a rookie's batting average during his first swing around the major league circuits has little relation to his average at the end of the season. By restricting the measure of batting success to the first month of play we prevent other samples of the rookie's batting prowess during later months from being selected. Hence, biased sampling is present.

A television rating service wants to find out what fraction of the television audience in a certain city are watching a particular program on Channel X at 6 p.m. of a certain day. Starting with a name selected at random in the phone book, the raters call every hundredth name and ask which station is being watched. Since the sample selected is limited to those who have both telephones and television sets, who have their television sets turned on, and who answer the telephone call, it is rather obvious that the sampling process is biased.

RANDOMNESS AND REPRESENTATIVENESS

Randomness refers to a method of selection in which each possible sample in a population has an equal chance of being drawn. It is certainly possible in finding the mean height of 9th grade students, for example, that the sample selected will contain only those who are below the mean in height. In this case the sample is certainly not representative of the whole population of 9th grade students. However, the theory of probability still applies as long as the method of selection gives each possible sample an equal chance of being selected.

Representativeness refers to agreement between the characteristics of a sample and the population from which it was drawn. Assuming a knowledge of the distribution of heights in a given population, for example, it is certainly possible to purposely select a sample whose frequency polygon

is similar to the frequency polygon of the population. In this case the sample could be representative of the population but probability theory could not be applied to the sample data, for the method of selection did not permit each possible sample an equal chance of appearing.

Let us examine some ways by which we might try to assure randomness in selecting a sample.

1. Suppose a door prize is to be given to one of the fifty couples who are attending a dance. As the couples enter each is given a ticket bearing a certain number. Toward the end of the evening one of the numbers is drawn from a container holding disks numbered from 1 to 50. If there has been a thorough mixing of the disks in the container, only chance should dictate the winner. This is *simple random sampling*. Each couple in the population had an equal and independent chance of winning. The same procedure could be used to draw samples of 2 or 3 or any number less than 50. However, the physical act of thoroughly shuffling large numbers of disks or cards or other objects and then drawing from them is much more difficult to do than is generally recognized. This, of course, is one reason some card players can seemingly read the cards. They remember the order of the cards which resulted from the play of the previous game and realize that, with the usual poor shuffling of the deck, the probability of approximately the same order is better than that of an arbitrarily different order.

2. You may think that by writing down a series of numbers which you select *just by chance* you would really get a random set. Unfortunately, this is not so, as most people will automatically tend to select numbers from the middle of the given group rather than at the end. An interesting experiment in a class might be to have each student write down ten numbers *at random*, these numbers to be chosen from the digits 1 to 10. You might expect that each number would appear about the same number of times. However, the results will surprise you. The chances are extremely high that the middle numbers 4, 5, 6, 7 will occur much more frequently than the extreme digits 1 and 10. For these reasons better ways of assuring randomness have been developed.

One such way is to use a table of random numbers.[15] Such a table of random numbers usually consists of rectangular arrays of 5-digit numbers which have been prepared by using mechanical and electronic methods which people accept as really producing a random series. Each number is located in a specific row and column of a specific page. The other numbers which are in the same row or column occur in such a sequence that there is practically no chance of predicting what these other numbers will be. By selecting a page at random, stabbing with a pencil at the selected page while the eyes are closed, to obtain an initial number,

flipping a coin to decide whether to proceed up or down the column or to the left or right in a row, one can be fairly certain that any bias in the chosen sequence of numbers is quite thoroughly eliminated.

Unfortunately, tables of random numbers are rather rare in school libraries and since they are fairly expensive, we look for a substitute. As a fair approximation to such a table of random numbers one may for classroom purposes use the columns of the fourth and fifth digits in a five place logarithm table. Suppose you want a sample of 10 children from a freshman class of 100. List the children alphabetically and number them in order. Then refer to the table of random numbers, and pick an arbitrary starting point. One might open the book casually and stab at the open page with a sharp pencil without looking. Starting with the number nearest to the pencil point take the next ten numbers in the columns ignoring duplications and calling 01, 02, 1, and 2, and so on with 00 to be 100. The children whose numbers come up in this way then constitute the sample. When we tried this method we obtained the ten numbers 11, 73, 34, 95, 56, 17, 78, 39, 00, and 61. Such a sampling procedure will be much more truly random than putting all the names in a hat, shuffling them and drawing ten names.*

Instead of taking a simple random sample of the whole population we may use other methods known as *stratified sampling* or *cluster sampling* or *systematic sampling*.

Stratified sampling is the process of dividing a population into subgroups and selecting a simple random sample from each of the subgroups. The total sample is the composite of the subgroup samples. The size of the subgroups in the sample is usually proportional to the fraction that the subgroups are of the total population. Sometimes this fraction is known; sometimes it has to be estimated. For example, suppose that in order to determine the college preferences of high school seniors we want to draw a sample but we feel that the sample should contain a number of seniors from each state which is proportional to the population of that state. Then from each state a simple random sample of predetermined size would be drawn. This would give a sample stratified according to area. The sample could have been stratified according to income tax brackets of parents or guardians. To do this we would draw from each income tax bracket a sample of size proportional to the number of people in that bracket. This kind of sample illustrates best what is known as a representative sample. Of course, what is a representative sample depends on the factors which are chosen to determine the representative-

* On closer inspection we note that this really gives a systematic sample, since the alternate numbers differ by a constant.

ness. It has been found very often that a stratified sample predicts the population characteristics better than a simple random sample.

An illustration of *cluster sampling*—any population can be considered as made up of groups of different sizes. For instance, suppose a public affairs organization wanted to get the reactions of high school students as to the question of going to college. Within a certain state a sampling of ten counties out of fifty might be taken by simple random sampling. Within each county five high schools might be selected by the same process. Neither the counties nor the schools might be of the same size. This would be an example of cluster sampling.

Systematic sampling. Suppose a survey is being made of telephone subscribers to find out which television program they are watching at a given day and hour. By simple random sampling a given page and line of the phone book is picked. Whether the left or right column would be used might be determined by flipping a coin. Then the corresponding telephone number would be called. Depending on the size of the listing every 50th or 100th or nth number after the first would be called until a certain number had been reached. The sample obtained would be called a systematic sample.

In our left-handed chair problem suppose we decide that from a population of 500,000 we want to draw a sample of 1000. We might do it by simple random sampling. Or we might decide on a stratified sample by selecting at random two students from each of the 500 schools in the system. A systematic sample might be taken by selecting ten schools at random and then taking every tenth name on the roster of those schools. Another way of getting a sample would be to select at random 25 schools and then again select in a random manner one room of forty students in each school.

Since in all the foregoing cases random sampling had been used somewhere in the process, the principles of probability could be applied. Hence, statistical inference of some sort would be possible in each of these cases. Of course, the conclusions that might be drawn would be limited by the nature of the population sampled in each instance.

SAMPLE SIZE

The size of the sample to select, as well as the method of sampling, is a problem of great importance in statistical inference. Consider the case of the boy who claimed that he could taste the difference between coke from a bottle and coke from a tap. We give him two glasses, one of each kind, and ask him to identify the two sources. By sheer luck he might make the correct selections. Is one sample of size 2 sufficient to either

confirm or deny his ability? If not, how large a sample should he used?

Suppose you are willing to admit the tester has the ability he claims if he succeeds in judging the contents of a number of glasses correctly 80 per cent (4 out of 5) of the time in the long run. Suppose you demand that he must not score below 75 per cent correct on any trial involving a certain number of glasses. How many glasses must he used at each tasting trial?

There are two kinds of errors of decision that might be made. One is that the taster may actually possess the ability but on the trial scores below 75 per cent; another is that he may not possess the ability but scores above 75 per cent. What can be done to minimize the risk of these errors?

What the statistician does is to select a *level of confidence*. A common one is the 95 per cent level of confidence. This means that if you conducted many trials of a specified sample size and made a decision about his 80 per cent ability each time, you would be correct, in the long run, 95 per cent of the time. If we tell the statistician that we are testing ability of an 80 per cent order, that we will accept on a single trial a score as low as 75 per cent, that we want to make decisions with a 95 per cent level of confidence, he will, in this kind of problem, be able to tell us the number of glasses we should use for our sample trial. The explanation of how he can arrive at that number is, in our opinion, beyond the ability of most high school mathematics classes.[16]

ACCUMULATION OF DATA

Once a problem and the population with which the problem is concerned have been well defined, and a sample is carefully selected in as unbiased a manner as possible, the characteristic data we want can be accumulated. This taking of data and recording it is something with which we are familiar to a greater or less extent. For example, consider our sample of 1000 obtained by selecting all 40 pupils from one room in each of 25 schools. Now the number of left-handed pupils per room is obtained merely by counting. Remember we count as left-handed for our purposes only those who actually write with their left hand and prefer left-handed chairs for writing. What shall we do with these numbers when we get them? Well, we have to record them somehow, and one obvious way would simply be to list the 25 numbers, each of which is somewhere between 0 and 40 inclusive. Such a simple listing is, however, fairly hard to analyze. Since one of the problems of statistics is to so order the primary data of the sample as to get as much information out of them as possible, we hunt for something better than a simple listing.

The numbers through which we record the data from our sample may

be obtained in two ways, depending on the type of problem. We may obtain them by *counting*, e.g., the number of left-handed students in a class, or by measuring, e.g., the length of life of an electric light bulb. Though we might want to check the accuracy of the count by having it repeated, there is not much disagreement as to how to make a count. About measurement, however, there is much to be said (Chapter 5). In the schools there is ample opportunity to discuss such ideas as the units of measurement, the accuracy, and the variability of the same measurement taken by different people, or by the same person at different times. However, in some cases even a single measurement must be considered as a sample of one of the many measurements that might be made. Such would be the case if an inspector were measuring the diameter of a piston in an airplane engine, or a physical education instructor were determining the weight of a boy in a gym class.

ORGANIZATION AND PRESENTATION OF DATA

We come now to the question of the organization of our data. As stated above, a mere listing of numbers is almost of no value. We must order them in some fashion before we can get much understanding of the information which is contained in them. This is usually done by making tables of the figures, which are often accompanied by graphs and diagrams. If there are large amounts of data, these must be summarized in convenient form. This is the sort of task which many junior high school classes would find fairly easy.

For example, suppose we have the following count of left-handed students in 25 classes of 40 pupils each:

5, 3, 4, 3, 2

2, 0, 2, 1, 2

3, 5, 2, 0, 6

4, 3, 6, 0, 1

2, 1, 3, 4, 2.

A useful summary of these data might be obtained by making a table of the number of classes having a certain number of left-handed pupils.

Number of left-handed pupils in class	0	1	2	3	4	5	6	7	⋯	40
Number of classes	3	3	7	5	3	2	2	0	⋯	0

From this table what is called a dot frequency diagram can be prepared which gives the data additional clarity (Fig. 2).

A still more useful method of presenting these data graphically is

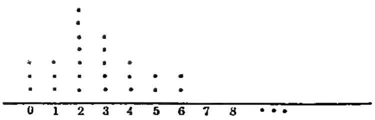


FIG. 2

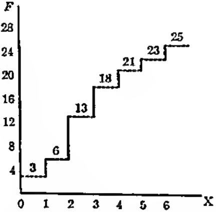


FIG. 3. F = number of classes having X or fewer left-handed pupils per class X = number of left-handed pupils per class.

through the use of the cumulative step graph illustrated in Figure 3. This graph is obtained by letting the ordinate F, for any given abscissa X, be the number of classes which have in them X or fewer left-handed pupils. Step graphs like this one may not be as familiar as bar graphs or straight line graphs, but there is no reason why children cannot learn to make them. There is a surprising amount of information which can be obtained from them.

For example, questions like the following could be asked: Which class contributed the greatest increase in the number of left-handed pupils? the least? In which X-intervals was there an increase of 2? of 3? Why is the largest frequency 25? What do the dots mean? What set of numbers is represented by X? by F? Where would you estimate the median to be? the 20th percentile? the 80th percentile? Which percentile is 3? 5? What is the probability that the number of left-handed pupils is less than 2 or greater than 4 in any one class?

Sometimes it is the case, especially when measurement data are involved, that we have a large number of scores each of which may occur only once or, as a statistician says, has a frequency equal to 1. In sum-

marizing such data we may find it more efficient to group the scores into classes.

For instance, the life in hours, to the nearest hour, of a sample of 40 light bulbs might be:

1043	976	989	1422	1053	891	1122	1236
1425	1213	1092	1321	943	985	1217	1212
1453	897	1213	972	1312	1114	1031	1172
1326	1317	987	1501	1211	1113	1215	1063
973	1421	1210	1231	1012	1311	1422	1342.

In order to simplify this array we can ask how many of these numbers lie between 875 and 925, how many between 925 and 975, and so on. In doing this if any score occurs which is exactly at a class endpoint, we must agree in advance to which class it should be assigned. In this instance we agree to place such a score in the lower class. For example, 1425 is assigned to the class with the midpoint 1400 rather than to the class with midpoint 1450. To do the grouping easily, a tally sheet may be set up as shown in Table 1. What we have done, for example, is to replace each of the three values 943, 972, and 973 in our original array by the midpoint of the class (950), do this for each class, and proceed to present this condensed picture of the original data.

A bar graph may be used to represent these data, as shown in Figure 4. This kind of bar graph is called a *frequency histogram*, a forbidding name

TABLE 1

Class	Midpoint	Tally	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
875–925	900	//	2	.05	2	.05
925–975	950	///	3	.08	5	.13
975–1025	1000	/~~///~~/	5	.13	10	.26
1025–1075	1050	////	5	.13	15	.39
1075–1125	1100	///	3	.05	18	.44
1125–1175	1150	/	1	.03	19	.47
1175–1225	1200	//// //	7	.18	26	.65
1225–1275	1250	//	2	.05	28	.70
1275–1325	1300	////	4	.10	32	.80
1325–1375	1350	//	2	.05	34	.85
1375–1425	1400	////	4	.10	38	.95
1425–1475	1450	/	1	.03	39	.98
1475–1525	1500	/	1	.03	40	1.00

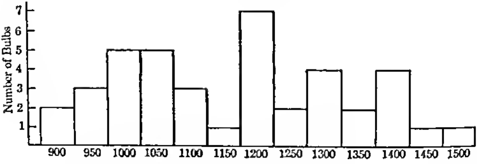


FIG. 4

for a familiar graph. The frequencies for each class are proportional to the areas of their respective rectangles.

When data are grouped in this way, it is customary to use a cumulative polygon graph, instead of the cumulative step graph, for further analysis of the data. This is done as follows: Take as abscissa the right hand endpoint of any class and determine the corresponding ordinate as the total number of members of all preceding classes. This is the number in the 6th column of Table 1. Thus, corresponding to 925, the endpoint of the first class, we plot 2, and corresponding to 1125 we locate 18, and so on. A starting point of 875 and 0 gives us the complete polygon when the plotted points are joined by straight line segments (Fig. 5).

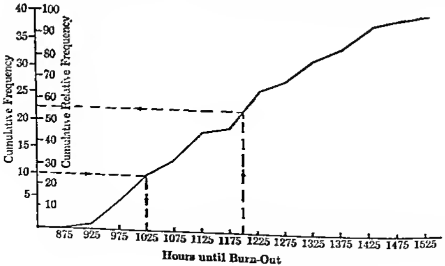


FIG. 5

weights of class members, the batting averages of the best 50 batters in the American or National Leagues at the end of the season, the total number of dots showing on each throw of three dice when they are thrown many times, the number of letters in each word of a page of a book, the distance students live from the school, the frequency of 0, 1, 2, $\cdots$ 9 in the last digit of street numbers, the number of calories in the food consumed by the class during a meal at home, and the number of children in each family of the class. The study of such frequency distributions will provide the students with procedures which are of considerable value in later work in statistics.

MEASURES OF CENTRAL TENDENCY

In addition to using the cumulative polygon to determine the median and the various percentiles of the set of observations, there are other ways of looking at the set to get more information. These are primarily concerned with two questions: What single score best represents the set of data as a whole? To what extent do the scores bunch together or scatter out?

There are at least three common answers to the first question; each of these has its advantages and disadvantages. The first is the median discussed above. This measure is easy to understand. It is easy to compute once the scores have been arranged in order of size. It is not usually influenced by a few scores which are much larger or much smaller than the other scores in the set of data.

To illustrate this last point, suppose you were considering the salaries paid in a certain factory. There were two janitors at \$2000 each, 10 workmen at \$4000, 2 foremen at \$5000, and the owner who received \$25,000. Obviously, the median of \$4000 gives a more representative measure of the salaries of the group than does the average salary which is \$5367.

The ordinary average whose value in the last example was \$5367 is called the *mean*. A more technical name for it is *arithmetic mean* to distinguish it from two other rarely used measures called the geometric mean and the harmonic mean. This arithmetic mean is influenced by extreme scores but it has some advantages which outweigh this fact. We list a couple.

The mean can be computed without arranging the scores in order of size. However, it does require more use of multiplication, addition, and division than does the median. Most important, if the population is separated into subgroups of known sizes and known means, the mean of the whole population can be determined from this information without

Bob's average is

$$\frac{70 + 70 + 73 + 62 + 86}{5} = \frac{361}{5} = 72.2\,.$$

and so forth for each boy. On the other hand, if the average grade in the mathematics class is wanted, we find that it is

$$\frac{80 + 86 + 92 + 71 + 63 + 50 + 80 + 75 + 77 + 81}{10} = \frac{755}{10} = 75.5\,.$$

What we did to obtain the average or mean of these mathematics grades was to divide the sum of them by the number of boys in the class. Suppose now we let n represent the number of boys in the class, and a, b, c, and so on stand for the mathematics grades of Allen, Bob, Charles, and so on. The class average in mathematics could then be symbolized as

$$\text{mean} = \frac{a + b + c + \cdots + j}{n}.$$

The $\cdots$ indicate the same ideas as the *and so on* written above, i.e., the grades of the other boys in the mathematics class.

We would want to use some other letters to represent grades in French or English. Would it not be simple to use the letter M to stand for a mathematics grade with an identifying subscript to indicate the boy involved? For example, M_a is Allen's mathematics grade, M_b Bob's, and so on. Even more simple is the idea of numbering the boys in some order. Thus, Let M_1 stand for Allen's grade, M_2 for Bob's, and so on. Then the mean could be written

$$\bar{M} = \frac{M_1 + M_2 + M_3 + \cdots + M_n}{n}.$$

In this formula $\bar{M}$ represents the mean of the mathematics grades. The mean of the grades in French would similarly be

$$\bar{F} = \frac{F_1 + F_2 + F_3 + \cdots + F_n}{n}.$$

However, the mathematician's passion for conciseness goes even further. In the case of the mathematics grades we observe that we added a group of terms of the form M_i, in which M indicates a mathematics grade and the subscript i refers to the boy whose grade was added. To give the command *sum all mathematics grades for all boys* we could more briefly say *sum all* M_i. As an abbreviation for the words *sum all* we will use the Greek capital S which is written $\sum$ and read as *sigma*. Hence,

adaptable to the expression of generalizations, and more tractable for the production of new relationships not previously perceived. In mathematics we must say exactly what we mean and only what we mean; we must be precise and concise. If we say too much, we are likely to mislead by causing the student to read into the excess language meanings we did not intend. If the kind of communication we use permits a symbol to have more than one meaning, then we are transmitting a message that is ambiguous.

As an illustration of the symbolism of statistics, consider the formal definition of the mean of a set of grouped measurements, such as the life of light bulbs we have used so often. This is

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}.$$

Certainly the first sight of this formula is enough to terrify anyone who does not know the meaning of its symbolic components. A very important contribution that teachers could make to an alleviation of this situation would be to introduce gradually some of the notations and symbolism which are really very useful even at early levels. We illustrate with a particular example.

SIMPLE MEAN

Suppose we want to compute average grades for individuals in a class and for a class as a whole. For illustration let us take a class of ten pupils each of whom takes English, Mathematics, History, French, and Art. We tabulate their grades as shown in Table 2. Allen's average is easily obtained as

$$\frac{69 + 68 + 85 + 73 + 80}{5} = \frac{375}{5} = 75\,.$$

TABLE 2

	Allen	Bob	Charles	Dick	Ed	Fred	George	Henry	Ike	Jack
Art	69	70	65	82	91	66	73	90	73	80
English	68	70	78	63	60	70	82	81	65	83
French	85	73	75	68	65	70	80	83	82	90
History	73	62	73	85	55	65	73	92	76	70
Math.	80	86	92	71	63	50	80	75	77	81

Suppose the different courses taken by the boys in our class above met for a different number of hours per week, e.g., Art for 2 hours, English and History for 4, and French and Mathematics for 5. Would an average computed by the method we have been using be a fair average? Most people would prefer a *weighted average*. This would be obtained by multiplying each grade by the number of hours the class meets, adding these products, and dividing by the total hours a boy has in class. Under this system Allen's mean would be determined by

$$\bar{a} = \frac{69 \times 2 + 68 \times 4 + 85 \times 5 + 73 \times 4 + 80 \times 5}{2 + 4 + 5 + 4 + 5}$$

$$= \frac{138 + 272 + 425 + 292 + 400}{20}$$

$$= \frac{1527}{20} = 76.4 \,.$$

If, as before, we let a_1 stand for Allen's Art grade, a_2 for the English grade, and f_1 and f_2 for the number of hours per week these courses meet, we may write

$$\bar{a} = \frac{f_1 a_1 + f_2 a_2 + f_3 a_3 + f_4 a_4 + f_5 a_5}{f_1 + f_2 + f_3 + f_4 + f_5} .$$

Using the Sigma notation this formula can be shortened to

$$\bar{a} = \frac{\sum_{i=1}^{5} f_i a_i}{\sum_{i=1}^{5} f_i} .$$

This is the frightening kind of formula we displayed originally. Perhaps our gradual approach to it has made it less fearsome.

MEAN OF A SET OF MEANS

The idea of an average of a set of averages is something which is vital in any statistical analysis. We could obtain the general class average, say $\bar{G}$, in either of two ways. We could find each boy's average $\bar{a}, \bar{b}, \bar{c}, \cdots \bar{j}$ from which we could obtain the grand averages,

$$\bar{G} = \frac{\bar{a} + \bar{b} + \cdots + \bar{j}}{10} .$$

We could also obtain $\bar{G}$ as the weighed average of the class average:

$$\bar{G} = \frac{2\bar{A} + 4\bar{E} + 4\bar{H} + 5\bar{M} + 5\bar{F}}{20} .$$

sum all M_i will be written as $\sum M_i$. To show exactly which boys' grades are being added, we will specify which values of i are involved by writing $\sum_{i=1}^{n} M_i$. This means that i takes all integral values from 1 to n inclusive. Again $\sum_{i=1}^{6} X_i$ means $X_3 + X_4 + X_5 + X_6$. Also, $\sum_{i=0}^{3} (\bar{X} - X_i)$ means $(\bar{X} - X_0) + (\bar{X} - X_1) + (\bar{X} - X_2) + (\bar{X} - X_3)$.

By using this kind of symbolism many sums can be written very concisely. Thus, the mean of the mathematics' grades is given by

$$\bar{M} = \frac{1}{n}\left[\sum_{i=1}^{n} M_i\right].$$

In like manner the mean of the French grades is

$$\bar{F} = \frac{1}{n}\sum_{i=1}^{n} F_i.$$

In this illustration we have omitted the brackets since they add nothing to the clarity of the formula. This last formula would be read: "F bar (or mean French grade) equals one over n times the sum of all F sub i where i goes from 1 to n." In like fashion we could express Allen's mean as

$$\bar{a} = \frac{a_1 + a_2 + a_3 + a_4 + a_5}{5}, \text{ or as } \bar{a} = \frac{1}{5}\sum_{i=1}^{5} a_i$$

and Bob's as

$$\bar{b} = \frac{1}{5}\sum_{i=1}^{5} b_i.$$

Suppose we want to find the class average for all grades. We can get it by adding the 50 grades and dividing the sum by 50. On the other hand we would obtain exactly the same result if we added the five subject means $\bar{A}$, $\bar{E}$, $\bar{F}$, $\bar{H}$, and $\bar{M}$, and divided by 5, or the 10 means of the boys $\bar{a}$, $\bar{b}$, $\bar{c}$, $\cdots$ $\bar{j}$, and divided by 10. It is just this property of the mean which, when proved in the general case, makes the mean very useful to the statistician in further mathematical developments. We note that if there were a different number of boys in the different classes we would need to use the idea of a weighted mean in order to get the grand average. This we develop in the next section.

WEIGHTED MEAN

We have not yet arrived at the general formula for the mean given at the beginning of our discussion of statistical symbolism. We will approach it through a simple example.

Each of the three rows of numbers has the same mean (5). However, the second row of numbers is grouped more closely about the mean than the other two rows. The numbers in the third row are more widely dispersed about 5 than those in the other two rows.

A first suggestion for finding a measure for the variability of scores about the mean might be to subtract each score, X_i, from the mean, $\bar{X}$, and sum these differences for each row of scores. In symbols what we will do is represented by

$$\sum_{i=1}^{n} (\bar{X} - X_i).$$

For the first row we obtain $(+4) + (+2) + 0 + (-2) + (-4) = 0$. For the second row we get 1, 1, 0, -1, and -1. This sum is also 0. For the third we have $(+4) + (+4) + 0 + (-4) + (-4) = 0$. Evidently, our first suggestion for a measure of scatter is useless, since it seems to always give us the same result, namely zero.

The students may think that the result is just an accident. They can obtain some interesting practice in dealing with signed numbers as they sum the deviations of scores from the means of such sets of scores. More important, the teacher has a good opportunity to discuss inductive thinking. (See Chapter 4.) How many sets of numbers must be studied in this way before some student will conclude that the sum of the deviations from their mean is always zero? What would constitute a satisfactory proof of such a theorem? As a matter of fact a proof of this statement, which we obtained inductively, can be given deductively. In the process valuable experience in the use of symbolism is provided. Briefly it goes like this:

$$\sum_{i=1}^{n} (\bar{X} - X_i) = \sum_{i=1}^{n} \bar{X} - \sum_{i=1}^{n} X_i = n\bar{X} - n\bar{X} = 0.$$

However, this proof requires some explanation and justification. First,

$$\sum_{i=1}^{n} (\bar{X} - X_i) = (\bar{X} - X_1) + (\bar{X} - X_2) + (\bar{X} - X_3) + \cdots + (\bar{X} - X_n).$$

By regrouping and using the associative and commutative laws of addition, we obtain

$$\sum_{i=1}^{n} (\bar{X} - X_i) = (\bar{X} + \bar{X} + \bar{X} + \cdots + \bar{X})$$
$$- (X_1 + X_2 + X_3 + \cdots + X_n).$$

The interesting thing about this is that we obtain the same result for the grand average no matter which way we compute it. Another simple example of a grand average would be the batting or fielding averages of a team from the averages of the individual members of the team.

A more complicated situation of the same general pattern might arise if a statistician were interested in a problem like the following: What is the average height of adult (over 20) males in the United States? He might take a sample of 100 men in New York City and find that their mean height is 68.3 inches. Another sample of 100 men in Philadelphia might yield a mean of 65.9 inches. He may proceed thus with 25 samples in 25 different cities, obtaining 25 sample means. Suppose he finds the mean of all of these sample means. Does he now have the true value of the mean height of *all* adult males in the United States? Obviously not. The problem confronting him is how close is he *likely* to be to this true value. Will it pay him in time and effort to seek to obtain a value closer to the true mean by taking more samples, or larger samples, or samples in a variety of places such as suburbs, rural communities, and other cities?

These problems are not at all simple and are bound up intimately with theories of probability, as the use of the word *likely* in the previous paragraph suggests. We will discuss these problems later. Actually we are suggesting again a fundamental problem of statistics, that of predicting the properties of a population from those of a sample. This prediction, however, will be valueless unless we can accompany it with some indication of the probability of the accuracy of the prediction. This reminds us of the confidence level we mentioned when discussing the trials of coke tasting ability.

VARIABILITY

A second question we promised to discuss was the bunching or scattering of scores. For example, every teacher knows that the means of two classes on a test may be the same but that the scores in one of the classes may be rather close together while in the other class there is a wide scattering. What we will seek now is a good method for measuring the average amount of scattering from the mean. As our intuitive notion of the scattering increases or decreases, so should our measure of this variation.

Consider, for example, the following three sets of observations:

$$1, 3, 5, 7, 9$$

$$4, 4, 5, 6, 6$$

$$1, 1, 5, 9, 9.$$

THE VARIANCE AND STANDARD DEVIATION

The property that made the $A.D.$ superior was the fact that absolute deviations are always positive. Since the square of either a positive or negative deviation would be a positive quantity, we can obtain the same advantage as the $A.D.$ by squaring the deviations from the mean before summing and averaging them. However, since this would magnify the measure of variation, we take the square root of the average of the sum of squares. The name given to this measure is the *standard deviation*. The average of the sum of squares of the deviations from the mean is called the *variance*. Hence, we can say that the standard deviation is the square root of the variance. In more advanced work in statistics the variance plays a much more important part than the standard deviation.

A general definition of the standard deviation is

$$s = \sqrt{\frac{1}{n}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]}.$$

In the symbolism we have adopted this formula can be expressed as:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\,.$$

We will now apply these formulas to determine the standard deviations of the three sets of numbers we used above.

$$s_1 = \sqrt{\frac{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{5}}$$

$$= \sqrt{\frac{16+4+0+4+16}{5}} = \sqrt{\frac{40}{5}} = \sqrt{8} = 2.8$$

$$s_2 = \sqrt{\frac{(4-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (6-5)^2}{5}}$$

$$= \sqrt{\frac{1+1+0+1+1}{5}} = \sqrt{\frac{4}{5}} = \sqrt{.80} = .89$$

$$s_3 = \sqrt{\frac{(1-5)^2 + (1-5)^2 + (5-5)^2 + (9-5)^2 + (9-5)^2}{5}}$$

$$= \sqrt{\frac{16+16+0+16+16}{5}} = \sqrt{\frac{64}{5}} = \sqrt{12.8} = 3.6$$

The standard deviation is the most widely used measure of the scatter or variability of a set of observations about their mean. One of the principal reasons for this is that it is defined in terms of mathematical opera-

Since the first parenthesis contains n terms, the sum $= n\bar{X}$. Since

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

we see that $X_1 + X_2 + X_3 + \cdots + X_n = n\bar{X}$. Therefore, the sum in the second parenthesis also equals $n\bar{X}$. By substitution, then,

$$\sum_{i=1}^{n} (\bar{X} - X_i) = n\bar{X} - n\bar{X} = 0\,.$$

A second suggestion for a measure of variability might be, "measure how far each X_i is from $\bar{X}$, ignoring the direction, and add up the results." In the three cases above this would give

$$(1)\quad 4 + 2 + 0 + 2 + 4 = 12$$

$$(2)\quad 1 + 1 + 0 + 1 + 1 = 4$$

$$(3)\quad 4 + 4 + 0 + 4 + 4 = 16.$$

This seems much better, since we agreed that the second set of observations is much less variable than either of the other two. Mathematically, this idea of the difference in size of two numbers without regard to the sign of the result is known as *the absolute value of the difference* and is denoted by $|X_i - \bar{X}|$. Note that it is customary to write X_i first; the same result would be obtained, however, if $\bar{X}$ had been given the first position. It is the sum

$$\sum_{i=1}^{n} |X_i - \bar{X}|$$

which we are considering here as a measure of variability.

For a large set of observations, widely scattered, the sum

$$\sum_{i=1}^{n} |X_i - \bar{X}|$$

might be awkward to handle. By dividing this quantity by n we obtain a measure of the average variation of each score from the mean. For each of the three sets of numbers in our example this *average absolute deviation from the mean* or *A.D.* equals 2.4, 0.8, and 3.2 respectively.

We can see that this measure gives a very sensible measure of average variation. Unfortunately, because the signs of the deviations from the mean are ignored, the *A.D.* cannot be treated mathematically in a very extensive way. As a result, statisticians have resorted to another measure of variation which has the good features of the *A.D.* but avoids the trouble with signs by a different device.

of problems which are susceptible to the methods we can develop easily. These methods are those that involve the use of the binomial theorem which is a familiar part of intermediate algebra. The development of such methods is known as building a mathematical model for the problem.

Two important uses of statistical inference are: (1) to test hypotheses, and (2) to obtain estimates of population measures from what we find out about a sample. In the first case we make a guess about a population measure, take a sample, observe the results of making that measure in the sample and then decide to accept or reject the guess. In the second case we start with a sample, observe the measures in the sample and then assert that the corresponding measure in the population lies between two numbers called the confidence limits. For example, (1) In a coming election for president of the junior class Jack's campaign manager may predict that his candidate will receive 60 per cent of the vote and Bill, his opponent, the remaining 40 per cent. The campaign manager's hypothesis is that Jack's per cent of the class vote will be 60. To test his hypothesis he questions a random sample of the class and finds that 55 per cent of those interviewed say that they favor Jack. As we will show later, with this information and the knowledge of a little mathematics he can determine at a certain level of confidence whether a sample of a certain size might give a value of 55 per cent if the true per cent were 60. (2) Instead of testing a hypothesis the campaign manager may want to obtain a lower and upper limit between which he can predict, with a high level of confidence that Jack's per cent of the vote will appear. To do this he takes a sample as before. Knowing the size of the sample and the per cent obtained from the sample, he can calculate a confidence interval for the true per cent by a method that gives correct decisions a high per cent of the time. We shall also illustrate this procedure later.

Another example of the first kind is the problem of the coke tester. The hypothesis might be that the claimant has no ability and the problem would be to decide how big a sample should be used and what results will enable the tester to say that his hypothesis is valid with a certain degree of confidence. The left-handed chair problem will give us another example of the second kind. From the results of the sample we want to predict the percentage of left-handed children in classrooms in the whole system within a certain margin and with a specified degree of confidence.

We have implied the need for a mathematical model in order to solve problems involving statistical inference. We have also indicated that we must restrict ourselves to a model associated with the binomial theorem.

tions which are easy to manipulate both arithmetically and algebraically. Another reason is that in many cases where large numbers of measurements are distributed approximately according to the so-called *normal* or bell-shaped curve the following results will hold, although we make no attempt to state precisely what conditions would have to be satisfied nor to prove the statement. Nevertheless, it is true that about 68 per cent of the observations will fall within one standard deviation on either side of the mean, 95 per cent within two standard deviations and 99.7 per cent (or nearly all) within three standard deviations.[17]

SUMMARY TO THIS POINT

We have, so far, outlined the importance of:

1. The careful statement of the problem which is going to be approached statistically.
2. Careful definition of the words used in the statement of the problem.
3. The necessity in many cases of drawing a sample.
4. Properties of samples and methods of drawing them to insure these properties.
5. The collection of data from the sample.
6. Methods of tabulating and organizing the data.
7. Symbolism and its advantages of precision and conciseness.
8. The use of the mean and the standard deviation to summarize the data.

We shall next consider the really fundamental problem—how to make inferences from the sample to the population. The solution of this problem will give us a sound basis for making the kind of decision the problem demands.

STATISTICAL INFERENCE

The problem of statistical inferences involves such questions as: Under what circumstances can we make inferences from the measurements in the population? When is a sample a good sample from which to make a prediction? How do you make such a prediction? Is the prediction a certainty and if not, how reliable is it?

We see that statistical inference which involves generalization from what is often inadequate data is a special case of inductive reasoning. It is subject to many difficulties and pitfalls and in order to study it we need to use the theory of probability. First, however, we should note that a complete discussion of the many types of problems which can be attacked by statistical methods and for which valid inferences can be made is impossible here since we must restrict ourselves to those kinds

any one element is independent of any other, i.e. whether a given child is left-handed or not does not depend on what is true of any other child in this room. This would in general be true unless there were a large number of identical twins or siblings of nearly the same age in the system and we would simply have to ignore this possibility in our study. (5) We are interested in the *total* numbers of *yes's*, i.e., not which individual child is left-handed but only how many left-handed ones there are in the room. We see that our problem satisfies all these conditions.

THE IMPORTANCE OF PROBABILITY

To construct our binomial model we are going to have to draw on some postulates of probability introduced in Chapter 6 so let us review them here.

1. We assume that if an experiment, which may or may not succeed, is tried repeatedly under exactly the same circumstances, there is a probability associated with the experiment which tells us how likely it is to succeed. This probability is a number between 0 and 1. We write

$$0 \leqq P(E) \leqq 1$$

where E refers to the occurrence of a specified event. Thus if E is the event of throwing an ace with a perfect die we assume $P(E) = 1/6$. Empirically we would say that this probability could be approximated by making n trials in which we obtained m successes and determining m/n. Of course n should be fairly large. Thus if a particular coin which is old and irregularly worn is tossed 1000 times and 435 heads are obtained we would say that the coin has a probability of approximately .435 of falling heads on any one toss, i.e., $P(E) \approx .435$. This is, of course, a different situation from that where we look at a coin before tossing it and after seeing that it looks symmetric and unbiased, *assign* a mathematical probability of $1/2$ or .5 for its falling heads. In the first case we assume that there is a fixed but unknown probability which we are trying to find. We attempt to determine it by repeated trials and a careful study of the result. In the second case we assign a fixed probability, in this case $1/2$. Another familiar example of this kind is that of a well-made die where *even* before it is cast we assign a probability of $1/6$ to the appearance of a particular face, say a 3. Probabilities assigned in advance of experiment are called *a priori probabilities*.

2. If the probability of one event is not influenced by the occurrence or nonoccurrence of another event, the two events are said to be independent in a probability sense. Thus two successive tosses of a coin, or two successive casts of a die, are assumed to be independent. The postu-

To use other than a binomial model would require a greater mathematical background than high school students possess. There are other statistical models or distribution functions, but a glance at the algebraic form of some of them, let alone an analysis of the assumptions which must be satisfied before they can be applied, is sufficient to reveal the necessity of our restriction. Some of these follow.

The Normal Distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

The Poisson Distribution:

$$f(x) = \frac{e^{-m} m^x}{x!}, x = 0, 1, 2, 3, \cdots$$

The Chi Square Distribution:

$$f(u) = \frac{1}{(\frac{1}{2}k - 1)!} \cdot \frac{1}{2^{\frac{k}{2}}} \cdot u^{\frac{k}{2}-1} \cdot e^{-\frac{u}{2}}$$

where

$$u = \sum_{i=1}^{k} \left(\frac{x_i - \mu}{\sigma_i}\right)^2 = \chi^2, \quad \text{and} \quad k = \text{the number of degrees of freedom.}$$

The concept of *model* is not a difficult one. For example, multiplication is a model for rapidly finding the sum of any number of identical addends. The Pythagorean theorem is a model for finding the length of any side of a right triangle, if two of the sides are known.

SPECIAL PROPERTIES OF A PROBLEM

We must next observe that when we restrict ourselves to a *binomial model* for the purposes of solving problems of statistical inference, we thereby limit ourselves to a certain type of problem. The nature of the restriction can be described by stating the five conditions that must be satisfied before the binomial model can be used. (1) There must be a certain fixed number of elements in the sample. Thus in the last example above, the classes have 40 children each. (2) For each element of the population from which the sample is drawn we must be able to decide *yes* or *no* as to whether the property exists, i.e., each child is left-handed (in the sense we defined above) or not. (3) The probability of *yes* must be the same for all elements, i.e., in this case we assume that there is a fixed probability of left-handedness in any individual. (4) The elements are independent in the sense that possession of the required property by

$$\begin{aligned} P(3 \text{ and } 2) &= P(R3 \text{ and } G2) + P(G3 \text{ and } R2) \\ &= P(R3) \cdot P(G2) + P(G3) \cdot P(R2) \\ &= \tfrac{1}{6} \cdot \tfrac{1}{6} + \tfrac{1}{6} \cdot \tfrac{1}{6} \\ &= \tfrac{1}{18}. \end{aligned}$$

These four probability postulates underlie the mathematical model we are building as the postulates of geometry underlie its structure. The applications of the model to the problems of statistics are like the application of geometrical theory to practical problems of measurement of physical space.

THE BINOMIAL MODEL

Before we study our main example let us consider a somewhat simpler one where the samples are of small size. We could start with a problem in coin tossing but it is a special case with a known probability of $\frac{1}{2}$ for heads on a true coin. Instead of a coin let us consider thumbtacks and investigate the determination of the probability of their falling point up when tossed. Since a thumbtack is far from symmetrical, we have no *a priori* basis for assigning probabilities to the possible events. We want first to analyse the situation theoretically and to make it reasonably easy we consider a case where $n = 3$, i.e., we toss three thumbtacks. Let $P(U)$ stand for the *probability of a tack falling point up.* If the tacks all are identical we may assume that this $P(U) = p$ is the same for each tack on each toss. We assume that there are only two possible positions of the tack, point up and point down. Let $P(D)$ mean *the probability of a tack falling point down.* Then by postulate 3, $P(D) = 1 - P(U) = 1 - p$ which we will call q for short. Finally, let us label our three tacks with red, white, and blue paint for identification purposes.

Now think of all of the possible events that might occur if we tossed the three tacks a very large number of times. Think of each toss as a random sample of all possible samples of size 3. On some of the tosses all three tacks might fall up. We will represent this outcome by the symbol UUU. There will also be cases in which we get outcomes like UUD where the symbols indicate in order the fall of the red, white, and blue tacks, red and white up and blue down. Besides these two there will be outcomes like UDU, and DDD and DDU.

By our second postulate the probability of $UUU = P(U) \cdot P(U) \cdot P(U) = p \cdot p \cdot p = p^3$. For the event UUD we obtain $p \cdot p \cdot q = p^2q$. For UDU we have $p \cdot q \cdot p = p^2q$ and for DUU we also get p^2q. Hence,

late we need is that the probability of two independent events both happening is the product of their probabilities. We write

If E_1 and E_2 are independent: $\quad P(E_1 \text{ and } E_2) = P(E_1) \cdot P(E_2)$.

Thus if a coin is tossed twice, the probability of throwing two heads is $(.5) \cdot (.5)$ or .25. If tossed three times the probability of getting three heads is $(.5)^3$ or .125. If a coin is tossed and at the same time a die is thrown, then the probability of throwing a head with the coin and simultaneously a 3 with the die is $(\frac{1}{2}) \cdot (\frac{1}{6}) = (\frac{1}{12})$. Again, if the probability of a vote for Jack in a certain election is $\frac{3}{5}$, then the probability that both of two arbitrarily chosen students will vote for him is $\frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25}$. Once more, if the probability is $\frac{4}{5}$ that a treatment will bring about a cure, then the probability that when two patients are given the treatment both will be cured is $\frac{16}{25}$.

3. Two events are mutually exclusive and exhaustive if both of them cannot happen but one or the other must happen. Thus heads and tails are mutually exclusive and exhaustive. In tossing a die the cast of a 2 and the cast of a 3 are exclusive but not exhaustive. But the cast of a 2 and failure to cast a 2 are exclusive and exhaustive. In this situation the sum of the probabilities is 1. We write:

If E_1 and E_2 are mutually exclusive and exhaustive $P(E_1) + P(E_2) = 1$.

A special case of this is when E_2 is the fact that E_1 does not happen. We can write $P(E) = 1 - P(\text{not } E)$. As an illustration the probability of throwing any face of a die except a 2 can be found as follows:

$$P\ (\text{not } 2) = 1 - P(2) = 1 - \tfrac{1}{6} = \tfrac{5}{6}.$$

4. If there are several mutually exclusive events, say $E_1, E_2, E_3, \cdots$, then the probability that one *or* the other of these events will occur is equal to the sum of their separate probabilities. We write: If E_1, E_2, E_3, ., ., ., are mutually exclusive, then

$$P(E_1 \text{ or } E_2 \text{ or } E_3 \text{ or } \cdots) = P(E_1) + P(E_2) + P(E_3) + \cdots.$$

Thus if a die is thrown once the probability of getting *either* a 3 or a 2 is $P(3 \text{ or } 2) = P(3) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Let us use these postulates to see how to find the probability of getting a 3 *and* a 2 if two dice are thrown simultaneously. Assuming one die is red and one is green in order to distinguish them, we find that since 3 *and* 2 means either [($R3$ and $G2$) or ($G3$ and $R2$)] and since the two dice are independent, we can say

is common in statistics to use the symbol $\binom{n}{x}$ instead of ${}_nC_x$ to represent the combination of n things taken x at a time. In this notation we would write

$$(p + q)^3 = \binom{3}{0}p^3 + \binom{3}{1}p^2q + \binom{3}{2}pq^2 + \binom{3}{3}q^3.$$

If we call falling point up a *success* we can summarize the results we have obtained in a table showing the probabilities of 0, 1, 2, or 3 successes in a toss of three tacks

x	0	1	2	3
$P(x)$	q^3	$3q^2p$	$3qp^2$	p^3.

Such a table is called a *binomial probability distribution* and is a special case of the many probability distributions which may occur. If we throw five tacks at a time the probabilities would be

x	0	1	2	3	4	5
$P(x)$	q^5	$5q^4p$	$10q^3p^2$	$10q^2p^3$	$5qp^4$	p^5.

Finally we can write the probability of getting exactly x successes in a binomial trial of size n in the functional form: $f(x) = \binom{n}{x}p^xq^{n-x}$. Here the domain of x is 0 and the positive integers from 1 to n while the domain of n is the positive integers.

To illustrate the use of the functional formula suppose we toss five tacks and assume that $P(U) = p = \frac{1}{4}$. To find the probability of getting 3 U's and 2 D's on a single toss we compute $f(3)$ when $n = 5$.

$$f(3) = \binom{5}{3} \cdot (\tfrac{1}{4})^3 \cdot (\tfrac{3}{4})^2 = 10 \cdot \tfrac{1}{64} \cdot \tfrac{9}{16} = .088$$

Similar computations show that when $n = 5, f(0) = .237, f(1) = .395, f(2) = .264, f(3) = .088, f(4) = .015$ and $f(5) = .001$. Of course these probabilities add up to 1 as they should.

Using this illustration and postulate 4 we can determine the probabilities of other interesting events. For instance, if x denotes as before the number of U's, we have $f(x < 2) = f(0 \text{ or } 1) = f(0) + f(1) = .237 + .395 = .632$. Also, $f(x \geq 3) = f(3 \text{ or } 4 \text{ or } 5) = f(3) + f(4) + f(5) = .088 + .015 + .001 = .104$. We think that this kind of application of the binomial model might be interesting and motivating to students.

The binomial $f(x)$ is a function in the same sense that the linear function $f(x) = mx + b$, with x real and m and b real constants, is a function. One principal difference is that the latter is a continuous function while the former is discrete. Functions like the binomial $f(x)$ are usually called distributions in statistics in the sense that they reveal what part of the total probability is associated with each value of x.

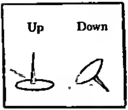


FIG. 6

by postulate 4 we find that the probability of two U's and one D equals $p^2q + p^2q + p^2q = 3\,p^2q$.

One U can occur either in UDD or DUD or DDU. By the same reasoning as for two U's the probability of each of these is $p \cdot q^2$. Hence the probability of the first or the second or the third tack falling up is $3\,p \cdot q^2$. Finally, DDD can occur in only that way. Its probability is q^3.

Again, by postulate 4 the probability of any one or the other of the four kinds of outcomes equals $p^3 + 3p^2q + 3pq^2 + q^3$. Notice that this equals $(p + q)^3$. This is the reason why our model is called a binomial model. If we had used four tacks, we would obtain

$$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

by the same kind of argument. Note that the coefficients in order from left to right equal

$$1, \quad \frac{4}{1}, \quad \frac{4 \times 3}{1 \times 2}, \quad \frac{4 \times 3 \times 2}{1 \times 2 \times 3}, \quad \frac{4 \times 3 \times 2 \times 1}{1 \times 2 \times 3 \times 4}.$$

These are also equal to the number of combinations of n things taken r at a time where $r = 0, 1, 2, 3$, and 4, respectively. Hence, they may be written as ${}_4C_0$, ${}_4C_1$, ${}_4C_2$, ${}_4C_3$, and ${}_4C_4$, respectively. That ${}_4C_0 = 1$ is justified by the theorem that ${}_nC_r = {}_nC_{n-r}$. Since ${}_4C_4 = 1$, then ${}_4C_{4-4} = {}_4C_0 = 1$. By similar reasoning if the sample size is n we find that

$$(p + q)^n = p^n + {}_nC_1 \cdot p^{n-1}q + {}_nC_2 \cdot p^{n-2}q^2 + \cdots {}_nC_{n-1} \cdot pq^{n-1} + q^n.$$

There are a couple of observations that may give us more insight into this binomial model. Since by postulate 3, $p + q = 1$, $(p + q)^3 = 1^3 = 1$. Hence the sum of the four terms of the expansion of $(p + q)^3$ also equals 1. This is reasonable because this expansion is the sum of the four probabilities which measure all of the possible outcomes in the toss of three tacks. In other words, one of the four (UUU, UUD, UDD, DDD) is *certain* to occur and the probability of an event which is certain to occur is 1. Another observation is that the number of different ways one of the four events can occur is given in terms of combinations. It

we should find out the probabilities associated with different numbers of successes from 0 to 20. Then we will have a better basis for making a decision.

What we need to do is to find the values of

$$f(x) = \binom{20}{x} \cdot (\tfrac{4}{5})^x \cdot (\tfrac{1}{5})^{20-x}$$

as x varies from 0 to 20. The computational work here would be very arduous. Fortunately, tables of the binomial distribution function are available.[18] In Table 3 we have included not only the values of $f(x)$ but also the cumulative probabilities represented by $F(x)$. To illustrate how the table should be read, suppose we want to know the probability of 18 successes in 20 trials. Find $x = 18$ in the first column and read .1369 in the second column. The probability of this event is, then, approximately .1369 or $\frac{1369}{10{,}000}$. To find the probability of 12 or fewer successes find $x = 12$ in the first column and read $F(x) = .0322$ in the third column. Care should be exercised in using the tables in the reference after using ours, as the entries are $1 - F(x)$ instead of either $f(x)$ or $F(x)$.

Our sample of 20 trials which showed 14 successes is only one of the

TABLE 3

x	$f(x)$	$F(x)$
0	.0000	.0000
1	.0000	.0000
2	.0000	.0000
3	.0000	.0000
4	.0000	.0000
5	.0000	.0000
6	.0000	.0000
7	.0000	.0000
8	.0001	.0001
9	.0005	.0006
10	.0020	.0026
11	.0074	.0100
12	.0222	.0322
13	.0545	.0867
14	.1091	.1958
15	.1746	.3704
16	.2182	.5886
17	.2054	.7940
18	.1369	.9309
19	.0576	.9885
20	.0115	1.0000

Another illustration is the application of the binomial model or distribution to batting average. Ty Cobb's batting average reached .400 several times in his career. Suppose in a given season this is true. We interpret this to mean that the probability of his getting a hit any time he is at bat is $\frac{2}{5}$. What is the probability that he gets exactly 2 hits on a day when he is at bat 5 times? We write

$$f(2) = \binom{5}{2} \cdot \left(\tfrac{2}{5}\right)^2\left(\tfrac{3}{5}\right)^3 = 10 \cdot \tfrac{4}{25} \cdot \tfrac{27}{125} = \tfrac{1080}{3125} = .346\,.$$

Hence, the chances are only about one in three that on a given day he will get just two hits. Of even more interest might be $f(0)$. This is

$$\binom{5}{0} \cdot \left(\tfrac{2}{5}\right)^0\left(\tfrac{3}{5}\right)^5 = \tfrac{243}{3125} = .077\,.$$

This means that the chances are about one in thirteen that even a .400 batter will go hitless in a given game when he comes to bat five times.

A different question from the one above is: What is the probability he gets *at least* two hits? The answer to this last question would be $f(2) + f(3) + f(4) + f(5)$ or more simply, $1 - f(0) - f(1)$. The reader should find it interesting to work out this result and compare it with the $f(2)$ found above.

TESTING HYPOTHESES WITH THE BINOMIAL MODEL

In an earlier part of this chapter we discussed in general terms a student's claim that he could distinguish between *tap coke* and *bottled coke* 80 per cent or $\frac{4}{5}$ of the time. Suppose we put one kind of *coke* in 10 glasses and the other kind in 10 other glasses and then have the student demonstrate his tasting ability. If in the 20 trials he is very unsuccessful in discriminating between the two liquids, we will reject the hypothesis that the student's probability of success is 80 per cent. If he is successful a little less than 16 times out of 20, we may either grant that he has the tasting ability he claims, or make no decision either way. Of course, we must admit that sometimes in samples of 20 trials he might score above or below 16 successes, since what he is claiming is an *average* ability of 80 per cent. We know from our study of the mean that there are bound to be scores above and below a mean, unless every score is the same.

This problem can be solved by using the binomial model. Since on each trial the student is either successful or not successful, we have the mutually exclusive case. The value of p is $\frac{4}{5}$ and of q is $\frac{1}{5}$. The value of n is 20 since that is the size of the sample.

Suppose in this experiment that the student scores 14 successes and 6 failures. On the basis of this single result can we say that the taster does not possess the 80 per cent ability? Before we answer this question,

decisions. The second kind of error is to accept the hypothesis when it is false. In the more advanced study of statistics the problem is to keep both of these errors as low as possible. Since each problem is different the expert statistician will find it necessary to set a different level of significance for different problems in order to minimize the two kinds of errors.

You may wonder why in our test of hypotheses we concentrate on rejection instead of acceptance. One reason is based on the idea that *one* exception to a general rule throws it out. It would be a tremendous task to take every instance and test for the correctness of the rule. For example, to test the misconception that the square of a number is always greater than the number, all we have to do is to show that the square of a number less than 1 is less than the number chosen. We do not have to run tests on all numbers.

ESTIMATING POPULATION MEASURES WITH THE BINOMIAL MODEL

In testing hypotheses we start with a guess about a population measure (parameter), take a sample, observe the results, and then make a decision to accept or reject the hypothesis. In estimating population parameters we start with a sample, observe the results, and then make a bet that the population parameter lies between two numbers called confidence limits.

For example, suppose we try to determine the unknown probability p of a tack falling point up. Obviously, if we throw a tack once, it may land U but this is no sure sign that it will always land U. So we throw a tack five times, or throw a set of five tacks once, to get a sample of size $n = 5$.

We find that 2 tacks land U. When a tack is thrown 10 times, point up may occur 5 times. When thrown 20 times, it may land point up 9 times. We recognize that the larger the sample, the better the chance we have of saying what p is with some degree of accuracy. However, we must also remember that any given sample may be one of those cases which are extremely unlikely but yet may perfectly well happen on occasion. Thus, with Ty Cobb and a p known to be .4 we found that in a sample of *size 5 the probability of getting no hits* at all was .077, which is not at all insignificant. We see that we would not be justified in going to a ball game and after watching a given batter go hitless say, "He must be a very bad batter because he didn't do a thing today."

To see how much difference the size of the sample can make in our problem suppose we first watch a .400 batter bat five times. In this

and $1 - F(x)$ to aid us in certain interpretations. For example, if $p = .04$, the probability that we will need at least 3 chairs is the probability that we will need 3 or more chairs. This is $f(x \geqq 3)$ which equals $1 -$ (the probability that we will need 0 or 1 or 2). Hence, $f(x \geqq 3) = 1 - F(2) = 1 - .788 = .212$. In general, the probability that we will need c or more chairs is given by $1 - F(c - 1)$.

On the basis of these figures and the analysis of them the architect is in a better position to make a decision. Of course, we must realize that these figures are not certainties. In the first place, we are only 95 per cent sure that the p we sought is between .04 and .06. Furthermore, if $p = .04$, the chances are 2 in a thousand that we will never need more than 6 left-handed chairs. If we were satisfied with chances of 2 in a hundred, we would need only 5 chairs. On the other hand, even if p is as high as .06 and we installed only 5 left-handed chairs, we would be on the safe side a little better than 9 times out of 10. We might well decide that this is good enough and tell the architect to go ahead and install 5 left-handed chairs in each classroom.

However, some critic may justifiably say, "p may well be only .04. If it is, then we are going to need even three of those left-handed chairs only 20 per cent of the time. Better put in only 2, and in the one room out of five where you need more, bring in some makeshift arrangement."

What shall we do? Well, at least our analysis has given us some basis for discussion. The decision for action can be made much more intelligently than if nothing were known, and the only idea prevalent was, "Let the left-handed children shift for themselves. There aren't enough of them to create a problem." We have now arrived at the long sought after result. We may be dissatisfied with the *iffy* quality of the proposed solution, but we must keep in mind that life is full of uncertainties. If statistical analysis can help us make better decisions in the face of these uncertainties, we and our students should learn how to put it to use.

CONCLUSION

The fact that statistical methods provide valid means of studying problems should not blind us to the concomitant fact that numerous errors can creep into statistical reasoning. Not for nothing has it been said that *anything can be proved by statistics*. It should also be realized that we have sketched here only the simplest kinds of situations, e.g., those that can be studied by the use of a binomial model.

Even in as simple a case as the kind we have outlined, many mistakes can be made, particularly if sampling errors are made. A sample may

be biased. It may not be large enough to yield results that are as precise as we need. Furthermore, we must remember that statistical results are uncertain as far as a single individual or a single case is concerned. Statistical results are valid only in the large. Again statistical measures are too often presented without clear definition of the terms used, or any statement about the reliability of the data offered. What does an average salary of $4150 in a given factory mean? Is it a mean, median, or a mode? Does it include the salaries of the executives, or only those of the union members?

The fallacies of statistics are common, and hard to detect. However, this fact should not keep us from realizing the great utility and far-reaching importance of the subject and its methods. The following quotation well summarizes our feeling about the place of statistics in education:

> Uncertainty dogs our every step . We must act on incomplete or unsure knowledge. .. The idea of sampling is an essential element for making sensible decisions, indeed it may be the basis of thought itself. We send out mental antennas to feel or taste the universe and from these samples which give us only partial information we learn to form sound judgments about the total populations they are supposed to represent.
>
> Our system of education tends to give children the impression that every question has a single answer. This is unfortunate because the problems they will encounter in later life will generally have an indefinite character. It seems important that during their years of schooling children should be trained to recognize degrees of uncertainty, to compare their private guesses and extrapolations with what actually takes place—in short, to interpret and become masters of their own uncertainties.[6]

See Chapter 11 for bibliographies and suggestions for the further study and use of materials in this chapter.

NOTES

1. NEWMAN, J. R. *The World of Mathematics*. New York: Simon & Schuster, Inc., 1956. Vol. 2, p. 1360.
2. COLLEGE ENTRANCE EXAMINATION BOARD. *Introductory Probability and Statistical Inference for Secondary Schools*. New York: the Board, 1957. p. 2.
3. WILKS, S. S. "Teaching Statistical Inference in Elementary Mathematics Curricula." *American Mathematical Monthly* 65: 145; March 1958
4. FISHER, RONALD. *The Design of Experiments*. Edinburgh: Oliver and Boyd, Ltd., 1949.
5. BROWNLEE, K. A. "Statistics of the 1954 Polio Vaccine Trials." *Journal of the American Statistical Association* 50: 1005–13; 1955.
6. WALLIS, W. A., and ROBERTS, H. V. *Statistics, a New Approach*. Glencoe, Ill.: The Free Press, 1956. Chapter 16.

7. Kemeny, J. G.; Snell, J. L.; and Thompson, G. L. *Introduction to Finite Mathematics*. New York: Prentice-Hall, Inc., 1957. Chapter 7.
8. Neyman, J., and Scott, E. "The Distribution of Galaxies." *The Scientific American* 195: 187; September 1956.
9. McCracken, D. D. "The Monte Carlo Method." *The Scientific American* 192: 90; May 1955.
10. Huff, D. *How To Lie with Statistics*. New York: W. W. Norton & Company, Inc., 1954.
11. Wallis, W. A., and Roberts, H. V. *Statistics, a New Approach*. Glencoe, Ill., The Free Press, 1956.
12. Siegel, S. *Non-Parametric Statistics for the Behavioral Sciences*. New York: Prentice-Hall, Inc.
13. Kolomgoroff, A. *Foundations of Probability*. New York: Chelsea Publishing Company, 1950.
14. *General Education in School and College*. Cambridge, Mass.: Harvard University Press, 1956. p. 56.
15. Interstate Commerce Commission. *Table of 105,000 Random Decimal Digits*. Washington, D.C.: Bureau of Transport Economics and Statistics. (Free)
16. Newman, J. R. *The World of Mathematics*. New York: Simon & Schuster, Inc., 1956. Vol. 3, p. 1512.
17. College Entrance Examination Board. *Introductory Probability and Statistical Inference for Secondary Schools*. New York: the Board, 1957. p. 30; Wilks, S. S. *Elementary Statistical Analysis*. Princeton, N. J.: Princeton University Press, 1951. p. 147.
18. College Entrance Examination Board. *Introductory Probability and Statistical Inference for Secondary Schools*. New York: the Board, 1957. Chapter 6, and Tables, p. 100, 101.
19. Wilks, S. S. *Elementary Statistical Analysis*. Princeton, N. J.: Princeton University Press, 1951. p. 201.
20. Cohen, John. "Subjective Probability." *The Scientific American* 197: 128–38; November 1957.

8

Language and Symbolism in Mathematics

ROBERT S. FOUCH AND EUGENE D. NICHOLS

"... and that shows that there are three hundred and sixty-four days when you might get un-birthday presents—"

"Certainly," said Alice.

"And only one for birthday presents, you know. There's glory for you!"

"I don't know what you mean by 'glory'," Alice said.

Humpty Dumpty smiled contemptuously. "Of course you don't—till I tell you. I meant there's a nice knockdown argument for you!"

"But 'glory' doesn't mean 'a nice knockdown argument'," Alice objected.

"When I use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."

"The question is," said Alice, "whether you can make words mean so many different things."

"The question is," said Humpty Dumpty, "which is to be master—that's all." —LEWIS CARROLL—

THINGS VS. NAMES OF THINGS

THE CENTRAL theme of this chapter is that of the difference between things and names of things and the implications of this difference for the student of mathematics. It seems that children and adults alike find it rather easy to distinguish between things and names of things in their everyday affairs. In mathematics, however, there is evidence to the effect that such distinctions are difficult to make. For example, the student who *cancels*

$$\frac{\cancel{4} + 3}{\cancel{4} + 10}$$

and, therefore, concludes that $\frac{7}{14} = \frac{3}{10}$ displays a symptom of such difficulty.

In the example above, *crossing out* the symbol '4' above the line and the symbol '4' below the line signifies subtracting the number 4 from

the number 7 in the one case, and subtracting the number 4 from the number 14 in the other case. Due to the nature of division, the result obtained by dividing 3 by 10 is not the same as that obtained when dividing 7 by 14. Thus, the *operation* of *crossing out* symbols must be checked by interpreting what is being done with the things (numbers) named by these symbols.

The student, using the same procedure in the case

$$\frac{\cancel{3} \times 2}{\cancel{3} \times 5}$$

and concluding that $6/15 = 2/5$ meets with the approval of the teacher. It might appear that using *cancellation* in the same way leads to an erroneous result in one case and to a correct result in another case.

In the last example, the *operation* of *crossing out* needs to be also interpreted in terms of operations with numbers. Crossing out the symbol '3' above the line means dividing 6 by 3. Similarly, crossing out the symbol '3' below the line means dividing 15 by 3. Due to the nature of division, the quotient is the same when both the dividend and the divisor are divided by the same number.

The kind of difficulty exemplified above in the case of arithmetic is even more intensified when students encounter algebra. For example, a student who *simplifies* as follows:

$$5x - x = 5$$

experiences the same difficulty. In interpreting the above to mean

Take away the symbol 'x' from the symbol '$5x$'

he fails to realize that there is a difference between *operating* on symbols and operating on things named by the symbols. Thus, the simplification of the example above might be done as follows:

When we wish to write *about a word or other symbol*, we shall put single quotation marks around that word or symbol.

Thus, we can now say correctly and without ambiguity that there is no student on the preceding page but 'student' occurs in several places on the page.

A few simple illustrations should prove to be of value at this point. Look at a children's puzzle which has gone the rounds of many a school.

When does half of twelve equal seven?
Answer: When it's written 'XII' and one takes the upper half.

Consider another puzzle of similar nature.

What is half of 18?
Answer: 10. It is obvious when you draw a horizontal line cutting '18' in half and take either one of the two halves.

Perhaps no one takes these puzzles seriously but we want to utilize them in illustrating the use of the single quotes we shall employ throughout this chapter.

It is easy to detect an analogy in a way in which one arrives at the erroneous answers in the three examples above:

It is not true that $5x - x = 5$, but it is true that the symbol 'x' removed from the symbol '$5x$' results in the symbol '5'.

It is not true that half of twelve equals seven, but it is true that taking the *upper half of the symbol 'XII' gives one the symbol 'VII'.*

It is not true that half of 18 equals 10, but it is true that taking either the upper or the lower half of the symbol '18' leaves one with the symbol '10'.

Now we are led to a topic which is of importance in the study of mathematics, viz., distinction between a *number* and *a name of a number*. To introduce this topic to a class of students, one may write symbols like these on the blackboard

$$\mathbf{3} \qquad 5$$

and then ask: "Which is larger?" A cautious student will insist that he cannot answer the question until he knows whether the question is about the marks he sees on the blackboard or about, as he may put it, "the numbers which are represented *by the marks*." The student, of course, is correct in not giving a simple answer.

We are making the following point. It is a simple matter to be able

which are intimately connected with the subject of the distinction between things and their names, in particular, between numbers and their names. The teacher who might think that this matter is of little consequence should attempt to uncover a good reason for students' frequent mistakes when they say that, for example, the number .0876 is larger than the number .7.

NUMBER—NUMERAL

The teacher in a little backwoods school was at the blackboard explaining arithmetic problems, and was delighted to see that the gangling lad, her dullest pupil, was giving slack-jawed attention. *Her happy thought was that, at last,* he was beginning to understand. So when she had finished, she said to him, "You were so interested, Cicero, that I'm certain you want to ask more questions."

"Yes'm," drawled Cicero. "I got one to ask—*where do them figures go when* you rub 'em out?"

In considering numbers, it is important to maintain the distinction between things and their names. Numbers, being abstractions, cannot be perceived by any of the five senses. It is impossible to distort the shape of a number, for it has no shape. It is impossible to shoot a hole through a number, for there is nothing physical to shoot a hole through. On the contrary, names of numbers can be seen, they can be erased, their shapes can be distorted, they can be moved, and many other sorts of physical actions can be performed on them.

Zero. Once awareness of this distinction is achieved, one must recognize that the statements to the effect that zero is not a number but *merely a place holder are incorrect. Such statements are open to criticism* on two counts. First, *zero is a number* (Chapter 2), for we write

$$5 + 0 = 5$$

when we want to express the fact that the number zero added to the number five equals five. Second, if we consider zero as a place holder, meaning that it actually *holds a place*, then we are referring to a mark on a piece of paper, for it would be absurd to talk about a number, which is an abstract entity, holding a place for some thing. The mark which we call a place holder is a name for the number zero.

To recognize this means also to recognize that devoting a great deal of attention to the numeral '0' as a place holder is *much-to-do-about-nothing*. Moreover, it is misleading, because of the failure to discuss other simple numerals like '1', '2', *and so on*, as place holders. In this sense, for example, in '7095' each of the numerals '7', '0', '9', and '5'

holds a place. '7' signifies the presence of 7 thousands, '0' the presence of no hundreds, '9' the presence of nine tens, and '5' the presence of five ones.

When looking at the statement

Ten divided by three is $\frac{10}{3}$

one sees the symbols 'ten', 'three', and '$^{10}/_3$'. They are names of three different numbers. But one also is able to recognize that the phrases

Ten divided by three

and

$$\frac{10}{3}$$

name the same number. (The reader by now has probably noticed that instead of including a phrase within single quotes, we occasionally display it.)

Our general point is that when one makes a statement about something, then it is normal for this statement to contain a name for this thing, rather than this thing itself. Or, expressed in other words, in order to mention something one uses a name of it.

It follows, then, that in making a statement about a number, it is natural that this statement should contain a name of the number and not the number itself. At different times, one may use different names of the same number. The choice of a name may be completely arbitrary or there may be a good reason for using one name in preference to some other name. For example, in books intended for German children the name '*zehn*' is much more appropriate than the name 'ten', although each names the same number. Or, ordinarily, the name '10' is considered to be simpler than the name '$^{160}/_{15} - {}^{10}/_{15}$'. Similarly, the name '9' is usually considered to be simpler than the name '2 + 7', although the latter may be simpler for some purposes. For example, when adding 9 to 28, a child will probably find it easier to do this:

$$28 + (2 + 7) = (28 + 2) + 7 = 30 + 7 = 37$$

than this:

$$28 + 9 = 37\,.$$

Thus, a number has many names. We choose the name which is simpler or more convenient for the purpose at hand.

Here are some of the names of the number 10:

$$30 - 20 \qquad 5 \times 2 \qquad \sqrt{100} \qquad \text{X}$$

$$\sqrt[3]{1000} \qquad 7 + 3 \qquad \textit{dix} \qquad \textit{zehn}\,.$$

It should be pointed out that $\sqrt{100}$ is the principal square root of 100, which is the number 10. The second square root of 100 is -10. Similarly, $\sqrt[3]{1000}$ is the principal cube root of 1000, which is 10. The other two cube roots of 1000 are complex numbers.

Some of the names of the number 3 are:

$$\text{three} \qquad \textit{drei} \qquad \sqrt{9} \qquad 7 + 4 - 8$$

$$27^{1/3} \qquad \frac{621}{207}\,.$$

Some of the names of the number $^{10}\!/_{3}$ are:

$$\text{ten-thirds} \qquad 10 \div 3 \qquad 3\tfrac{1}{3} \qquad \frac{170}{51}$$

$$\sqrt{\frac{100}{9}} \qquad \frac{60}{15} - \frac{10}{15}\,.$$

Since the names in the last line above are names of the same number, we may write, for example, the following:

$$3\tfrac{1}{3} = \frac{170}{51}\,.$$

This statement implies that '$3\frac{1}{3}$' and '$^{170}\!/_{51}$' are names of the same number. Similarly,

$$\text{ten} = 10$$

implies that 'ten' and '10' are names of the same number. And

$$\text{ten} = \text{the sum of five and five}$$

implies that 'ten' and 'the sum of five and five' are names of the same number.

Correspondingly, the sign '$\neq$', which is read *is not equal to* is interpreted to mean that two names connected by it name two different entities. For example,

$$\sqrt{100} \neq \frac{10}{3}$$

means: '$\sqrt{100}$' and '$^{10}\!/_{3}$' are names of two different numbers.

The reader is referred to the section on equality toward the end of this chapter for a more detailed discussion of the concept of equality.

Fractions. If we now turn our discussion to numbers which are fractions, we first realize that the word 'fraction' may at times be ambiguous. For the sake of clarity, we will use here the phrase 'fractional number' for the number and the phrase 'fractional numeral' for the symbol for such a number. For example, the fractional numerals '$\frac{1}{2}$', '$\frac{2}{4}$', '$\frac{3}{6}$', and '$\frac{4}{8}$' are different names of the same fractional number. On the other hand, the fractional numeral '$\frac{1}{3}$' is not a name of the same rational number as the fractional numeral '$\frac{1}{2}$'.

Thus, *a fractional numeral is a symbol naming a fractional number.* We use the phrase 'fractional number' to be synonymous with 'rational number'. Since in common usage the word 'fraction' is used to refer to a number, we may abbreviate and also use 'fraction' to be synonymous with 'fractional number' or 'rational number'. You will recall the discussion in Chapter 2 in which a rational number was defined to be an equivalence class whose elements are number pairs (page 43). An example of such an equivalence class may be

$$\{(1,2), (2,4), (3,6), (4,8), \cdots\}.$$

The fractional numeral '$\frac{1}{2}$' is one of the many names for this equivalence class. The class itself is a rational number.

The discussion of the distinction between names of things and the things themselves, in particular, the distinction between a fractional numeral and a fractional number, has many implications for teachers. Let us examine a number of statements currently occurring in arithmetic textbooks and see how they should be modified in the light of the preceding discussion.

(1) Parts of whole things, such as $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{3}{4}$, are called fractions, or fractional parts.

In the statement above, it appears that the phrases 'fractions' and 'fractional parts' are used as synonyms. Also, fractions are considered to be parts of something, e.g., parts of objects, like parts of an apple. At the same time, it is said that $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{3}{4}$ are fractions; thus, they are parts of objects. Whether one considers fractions to be parts of either common objects or mathematical objects, i.e., numbers, one should experience a feeling of uneasiness about it. For, first, no number can be a part of a common object, and, second, a rational number should not be considered a part of a number.

A clearer statement saying what was apparently intended by the

statement above would be:

(1′) Numbers such as 1/2, 1/3, and 3/4 are called rational numbers; more explicitly, a number is a rational number if it is the quotient of two whole numbers. The symbols '1/2', '1/3', and '3/4' are examples of fractional numerals.

Let us consider another statement:

(2) The figure written below the line is called the denominator of the fraction and the figure written above the line is called the numerator of the fraction.

According to this statement, the numerator and denominator of a fractional numeral are chosen to be symbols. One should not quarrel with this choice, except for the fact that those who make this choice go on and perform operations on numerators and denominators, that is, on symbols, instead of on numbers named by the symbols. Since it is quite convenient to speak of, for example, multiplication of a numerator by a number, it is easy to agree that the numerator be the number named by the top part of a fractional numeral and the denominator the number named by the bottom part. Thus, the statement above would be written as:

(2′) A fractional numeral consists of two parts. The top part names the numerator, and the bottom part names the denominator.

Now, since a numerator and denominator are numbers, we can freely speak about multiplication of the numerator and denominator by a number.

In the light of the discussion above, the following commonly used definition:

(3) Fractions which have the same denominator, such as 2/5, and 3/5, are called like fractions.

should read:

(3′) Fractional numerals with the same denominators are called 'like fractional numerals'.

Thus *likeness* is not a property of numbers; it is a property of numerals. For example, the fractional numerals '3/4' and '$\frac{7}{5-1}$' are like fractional numerals because their bottom parts name the same number, i.e., they have the same denominator.

In order to test whether or not this distinction between the number

and the numeral has been successfully taught, we would suggest trying on students exercises like the following. For each statement below, tell whether it is true or false. State the reason for your answer.

1. 2 > 3.
2. '45' consists of two numbers.
3. '5' is *smaller than* '5'.
4. 5 is larger than 5.
5. 'Five' has four letters.
6. 100 is made up of '1', '0', and '0'.
7. Johnny can erase '5' in 35.
8. In writing '17' we write 1 first, then 7.
9. In '56', '5' is bigger than '6'.
10. Four has four letters.
11. If '5' is bigger than '9' in '59', then 5 > 9.
12. '0.00005' is larger than '0.5'.
13. 0.5 is larger than 0.00005.
14. The fractional numeral $\frac{'2'}{3}$ is smaller than the fractional numeral '$\frac{2}{3}$'.
15. Numbers can be found on this page.
16. 'Numbers' can be found on this page.
17. '$\frac{5}{7}$' and '$\frac{5}{10-3}$' are like fractional numerals.
18. '$\frac{3}{5} \neq \frac{6}{10}$' means that '$\frac{3}{5}$' and '$\frac{6}{10}$' are names for two different numbers.
19. It is true that 7 < 8.
20. There are nine letters in 'one letter'.

Answers:

1. False. The number two is not greater than the number three.
2. False. No symbol consists of numbers.
3. True. It is easy to *see* that the first symbol is smaller than the second symbol.
4. False. The number five is not larger than the number five.
5. True. One verifies this by simply counting the letters in the word 'five'.
6. False. *No number is made up of symbols.*
7. False. No symbol can be erased in a number, because no number is made up of symbols.

8. False. It is impossible to write numbers.
9. True. The symbol '5' is bigger than the symbol '6'.
10. False. Numbers do not have letters.
11. False. It is true that the symbol '5' is bigger than the symbol '9' in the symbol '59'. It is false that $5 > 9$. Thus the total statement is false.
12. True. It is easy to *see* that the first symbol is larger than the second.
13. True. The number 0.5 is larger than the number 0.00005.
14. False. It is easy to *see* that the first symbol is larger than the second.
15. False. It is impossible for numbers to appear on a page.
16. True. The word 'numbers' is capable of being found on a page. Furthermore, it is found in problem 15.
17. True. The two fractional numerals have the same denominator, since '7' and '10 − 3' name the same number.
18. True. The statement is about the *meaning* of the statement:

$$\frac{1}{2} \neq \frac{5}{10}.$$

It is true that the meaning of this statement is that '$\frac{1}{2}$' and '$\frac{5}{10}$' are names for two different numbers. However, the statement

$$\frac{1}{2} \neq \frac{5}{10}$$

itself is a false statement, because we know that '$1/2$' and '$5/10$' *are* two names of the same number.
19. True. The statement is about the numbers seven and eight, and we know that the number seven is less than the number eight.
20. True. The statement is about the phrase 'one letter'. By counting the letters in this phrase, we can verify the statement.

Further importance of distinguishing between numbers and their names, especially in statements where reference is made interchangeably to numerals and numbers, can be illustrated by the following example.

Suppose we assume, as is commonly done in mathematics textbooks, that the statement

$$\frac{6}{8} = \frac{9}{12}$$

implies that '$6/8$' can be replaced by '$9/12$' in any statement. This is in accordance with the axiom which is usually stated in some such ambiguous fashion as "Equals may be substituted for equals in any expression without changing the value of the expression."

Now the statement

The denominator of '$\frac{6}{12}$' is exactly divisible by 3

becomes

The denominator of '$\frac{4}{8}$' is exactly divisible by 3

when '6/12' is replaced by '4/8'. The last statement is clearly false.

To avoid this difficulty, it is essential to note first that the statement

The denominator of '$\frac{6}{12}$' is exactly divisible by 3

is a statement about the number 12. Obviously, what is true of the number 12 does not have to be true of the number 8. Similarly, what is true of the symbol '$\frac{6}{12}$' is not necessarily true of the symbol '$\frac{4}{8}$', because they are different symbols. For example, the statement

'4' is a part of '$\frac{4}{8}$'

describes something which is a property of the symbol '4/8', but is *not* a property of the symbol '6/12'.

The fact that a given number is a single abstract entity suggests that the properties of a number are not subject to change determined by the form of a name of the number. Consider, for example, the number 7. The symbol '7' is one of its many names. It is a Hindu symbol in our Arabic numeration system to the base ten. One of the properties of the number 7 is that it is a prime number, i.e., it is divisible by only 1 and 7.

Numeration. A word of explanation of the use of the phrase 'numeration system' as contrasted with 'number system' is of importance here. The phrase 'numeration system' refers to a system of *writing* numerals. Thus, we may have a numeration system to the base ten employing ten symbols; or a numeration system to the base two employing two symbols, and so on. For example, '15' in the numeration system to the base ten, '30' in the numeration system to the base five, and '1111' in the numeration system to the base two name the same number.

The phrase 'number system', on the other hand, refers to a system of classifying *numbers*. Thus, we may speak of natural numbers, or integers, or rational numbers, and so on.

Consider now a name for the number seven in the binary numeration system, i.e., numeration system to the base two. The question is: "Is the number whose name in the binary system is '111' a prime number?" Certainly so. Primeness is a property of a number and therefore does not depend on the particular name we happen to choose for this num-

ber just as whiteness of snow does not depend on whether it is called 'snow', 'schnee', 'neige', or 'nieve'.

Let us now consider another example. Frequently the following question is asked: "Is π an exact number?" First, we hasten to point out that such a question should not be asked, because it suggests that numbers can be classified into exact numbers and inexact numbers. But there are no numbers which are inexact.

Our concern here is, of course, with what is commonly called 'decimal approximation'. We know of many numbers which differ from the number π by a very small number. Thus, 3.14 is an exact number. It differs from the number π by a small number; therefore, we may refer to it as being *an approximation of* π. We also know that no matter how many decimal places we use in a decimal, we will never be able to obtain a name for the number π. Furthermore, we know that the quotient of the circumference by the length of the diameter of any circle is a number. This number has no name in our Hindu-Arabic numeral system. It was given a name 'π'.

Perhaps it is a case that the habit of naming numbers using decimal numerals is so strong that we are tempted to believe that if there is no name of this kind, then there is no number. A good analogy for teaching purposes is the foreign language word which has no English translation or the fact that there is no Polynesian word for snow. In the former case, the absence of the English translation does not imply that there does not exist an object named by the word in the foreign language. In the latter case, the absence of a word in the Polynesian language for snow does not imply nonexistence of snow.

This suggests another example, at the high school level, with a reverse twist: namely, the symbol, or more properly, the mark '∞'. This mark names no number. In the simple theory of limits, it is pointed out that the function described by '$y = 1/x^2$' does not have a limit as x approaches zero. It is not only true that, upon replacement of 'x' in '$1/x^2$' by '0', we obtain '$1/0$', which is a meaningless symbol and, therefore, does not name a number, but also, numbers named by '$1/x^2$' when 'x' is replaced by names of small nonzero numbers are very large numbers. This fact is often described by saying that "as x approaches zero $\frac{1}{x^2}$ has no limit because it increases without bound." This statement is written thus in symbols:

$$\lim_{x \to 0} \frac{1}{x^2} = \infty.$$

So far all is well. But the student sees the mark '∞' appearing in precisely the same kind of situations in which numerals are used. He is soon treated to the sight of such expressions as '∞/∞', '$\infty - \infty$', '1^∞', and so on. Since the mark '∞' appears in contexts in which the student is accustomed to seeing symbols such as '5', '1', and so on, he eventually ends up by believing that '∞' also names a number. The truth here is that contexts in which '∞' occurs, although formally similar to some in which numerals occur, actually refer to facts concerning limits, and so have quite different meaning from those of the similar expressions containing numerals. The mark '∞' in the context of limits does not name a number. Thus, we have, in this case, what looks like a name of some thing but no thing to be named.

Our remarks here need to be interpreted in relation to the real number system. In this chapter we are not concerned with the extensions of the real number system, called 'transfinite numbers'. Our remarks here are only applicable to the simple theory of limits and to the system of real numbers as treated in the introductory calculus texts.

It is now time to bring out in the open a concept which has already been much used in the above. This is the matter of the symbol and its referent. We begin with a simple example. We all know that there once existed a flesh and blood person whose name was 'Euclid'. The name 'Euclid' is not flesh and blood but merely printer's ink in a certain pattern on this page; it is a symbol and the man is the *referent* of the symbol. In the same way, 'United States of America' is an approximately two-inch long set of black marks and a symbol, but the referent of this symbol is a very large nation of many millions of people and several millions of square miles of land.

As long as the referent of a symbol is a physical entity, it seems easy to avoid confusion—one can often rely on an index finger and point to the referent and point to the symbol naming this referent. In mathematics, as well as in many other areas of intellectual endeavor, troubles arise when the referent of a symbol cannot be pointed to or even imagined visually with any accuracy because that referent is abstract. Numbers are such things. The symbol, however, is always tangible, perceivable either by the eye or the ear, and there is therefore a strong tendency to think about the symbol in place of its referent. This is the source of such practices as *juggling numbers* and an explanation of why transposition is such a favored method of solving equations.

It is not our intention to condemn such practices as transposition and cancelling. We do believe, however, that students need to understand that transposition is a process of manipulating symbols and *not* of operating with numbers. Thus, a typical statement of a transposition rule to

the effect that a number can be moved to the other side of the equation if its sign is changed is not correct.

We feel it important that the student who learns to do this:

$$3x + 5 = 2$$

$$3x + 5 - 2 = 0$$

understands what is being done. We further believe that the use of the word 'transposition' contributes to neither understanding of the process itself nor proficiency in solving equations.

If we were to make a practical suggestion to teachers, it would be this:

> Let the students observe how equations are solved. Let them see what happens to the numerals and variables as they are *moved* from one side of an equation to the other, and do not attempt, at least in the early stages of work with equations, to verbalize the process or ask students to do it. To find out whether or not the students know what takes place, let them work equations and observe their work to see whether they are doing it correctly.

For more detailed treatment of solution of equations we refer the reader to the section on equality toward the end of this chapter.

Cancelling is an example of another process of manipulating symbols and *not* of an operation with numbers. Unless a student understands this clearly, he will make errors of the type shown in an example at the beginning of this chapter. Thus, he will cancel:

$$\frac{\cancel{2} + 3}{\cancel{2} + 6}$$

and conclude erroneously that $\frac{5}{8} = \frac{3}{6}$. Errors of the same type committed where variables are involved such as:

$$\frac{\cancel{x} + 5}{\cancel{x} + 8} = \frac{5}{8}$$

may not be as easily detected. They stem basically from the confusion about the meaning of 'variable'. This topic is discussed later in the chapter.

MATHEMATICAL GRAMMAR

> The study of grammar, in my opinion, is capable of throwing far more light on philosophical questions than is commonly supposed by philosophers. Although a grammatical distinction cannot be uncritically assumed to correspond to a genuine philosophical difference, yet the one is *prima facie* evidence of the other, and may often be most usefully employed as a source of discovery.—BERTRAND RUSSELL

Grammar is concerned with the structure of language. The person who has had formal training in one or more foreign languages is aware that each language has its own grammar, that different languages have different grammars. So it is with the *language of mathematics*. In this section we shall attempt to develop a very brief mathematical grammar with certain parallels to conventional English grammar.

First, however, we want to point out that our choice of the phrase 'the language *of* mathematics' is not without purpose. We believe that one should not speak of mathematics as a language because it is not a language in the same sense as Chinese is. Mathematics is a science studying a variety of things such as numbers, lines, and so on. A mathematical law such as the commutative property of addition can be expressed in many different languages. For example:

> The sum of one number and another number is equal to the sum of the second and the first.

is one English expression of this law. There are other English wordings and translations in most foreign languages. An expression of this in the language of mathematics might be:

$$x + y = y + x.$$

Our point is that the mathematics is the same in each case but the language is different. The Arabs were able to do a surprising amount of algebra without special symbols. However, contemporary mathematics, within certain limits, does have a fairly uniform method of expression and this can, roughly speaking, be called 'the language of mathematics'.

It should be understood that in this section we are discussing that part of mathematics which is expressed, not in ordinary English words which must follow conventional English grammar, but in that abbreviated symbolism which is special to mathematics.

It is usual in grammar first to discuss the various parts of speech, e.g., nouns, pronouns, verbs, adjectives, and so on. For most languages, such analysis is fairly complex and the different word forms are very numerous; this is partly due to the natural evolution of a language and partly to the great variety of concepts a language must express. On the other hand, the language of mathematics has been developed, at least in part, in a conscious, artificial way and also is used to express far fewer and probably far simpler concepts. As a result, we can identify only three rather clearly defined parts of mathematical language.

The first is almost completely comparable to the ordinary noun or pronoun, which is commonly said to be a name of a person, place, or

thing. While we have no persons to talk about in mathematical discourse, we do have places, such as the points of geometry, and things, such as numbers, sets, and so on. The phrase 'mathematical individuals' is sometimes used for these places and things. So, our first part of mathematical speech is the symbol for an individual; the simplest examples are the numerals such as '0', '1', and so on, and the letters 'x', 'y', 'A', 'B', and so on, used as variables in algebra and geometry.

The second part of mathematical language, the relation symbol, is very similar to the ordinary verb, which is said to express an action or a relation. Mathematics is not concerned with actions but there are very many relations of concern to the mathematician. One of the first mathematical things the young child learns, $1 + 1 = 2$, involves the relation of equality and this relation is never discarded even in the most advanced mathematics, even though many other relations are introduced. Other examples are the relation of being larger (expressed by '$>$'), the congruence relation (expressed by '$\cong$'), and the membership relation (expressed by 'ϵ'). Just as the English teacher insists that a correct sentence must have a verb, so the mathematics teacher must insist that a correct mathematical sentence must have a relation symbol.

It is difficult to compare our third part of mathematical language, symbols for operations and functions, with anything in traditional grammar. We are concerned here with such symbols as '$+$', '$-$', '$\times$', '$\div$', '$\sqrt{\ }$', 'sin', 'tan', '$\int$', and so on. It seems best simply to examine the way in which they are used. For example, '$+$' is a symbol for a binary operation, that is, if a numeral is written on each side of the sign, the resulting compound symbol is a name of some number. Some insight may be gained by translating '$2 + 3$' as 'the sum of 2 and 3' and comparing it with the ordinary English phrase 'the elder son of Adam and Eve'; conceptually, there is very little difference but the usual symbolism hides the similarity. The difference between a binary operation (as symbolized by '$x + y$') and a function of two variables (as symbolized when writing '$f(x, y) = x^2 - y^2$') is largely symbolic; certain functions are so basic or so much used that they have acquired a special symbolism and the title of an operation. There are also important symbols which combine with a single individual-symbol to produce a compound name of another individual; examples are '$\sqrt{\ }$', 'tan'. Here, we have singulary* operations or functions of one variable. In each case, if an individual-symbol is put in the proper place, the result is a compound symbol naming an

* Unfortunately, there is no agreement in terminology for this kind of operation; the words 'unitary', 'monary' are also used.

individual; for example, '$\sqrt{4}$' is a compound name for the number 2. A worthwhile comparison is to the English phrase 'the mother of Napoleon'. We shall call symbols of both operation and function 'operator symbols'.

The use in algebra of the symbols '+' and '−' in front of a numeral is an example of a misleading use of symbolism. The symbols '+5' and '−5' are frequently read 'positive five' and 'negative five', creating the impression that '+' and '−' appear here in the role of adjectives rather than merely as a part of the name of a number. *This is a source of confusion to those students who think that* $-x$ is a negative number. It seems preferable not to use the '+' sign at all, except possibly for emphasis in some cases, and to view the '−' sign as symbolizing a singulary operation. Thus '$-x$' should be read 'negative of x'. It is unfortunate, for some reasons, that '−' symbolizes both a singulary and a binary operation; for the mature person, this is a convenience because of the close relation between the operations, but for the immature, it can be a danger which the teacher must be prepared to handle properly. Similar remarks hold for the word 'negative' and the phrase 'negative of'.

Each of the three kinds of symbols (individual, relation, and operation) may be classified in either one of the two following categories: constants and variables. Thus, we shall speak of individual constants or individual variables, relation constants or relation variables, and operator constants or operator variables. Table 1 shows these categories more clearly, together with some examples in each category.

Although relation variables and operator variables are at present used mainly only at a rather advanced level, there is considerable opportunity for their advantageous use much earlier. For example, in 9th grade algebra, after one has explained that addition and multiplication are commutative, one can use an operator variable '$*$' to make a general definition of commutativity:

> A binary operation symbolized by '$*$' is commutative if and only if $a * b = b * a$, for all a and b.

In a similar way, a better understanding of algebraic and geometric relations may be produced through a general discussion of reflexive, symmetric, and transitive properties. For example:

> A relation symbolized by 'R' is *symmetric* if and only if: if aRb, then bRa, for all a and b.

There is some value in further analysis of the concept of compound names. Suppose that we use the phrase 'atomic constant' for an individ-

TABLE 1

	Constants	Variables
Individuals (symbols for things)	2 the point (0, 0), ∅ (the null set)	x (The range of 'x' may be the set of complex numbers.) A (The range of 'A' may be the set of points in a plane.) B (The range of 'B' may be a specified collection of sets.)
Operators (symbols for functions and operations)	$+, -, \div, \times$	(Generally found only in advanced mathematics.) o (The range of 'o' may be the entities named by: '+', '−', '×', and '÷'.)
	sin	f (The range of 'f' may be the entities named by: 'sin', 'cos', 'tan', 'cot', and 'log'.)
Relation symbols	$=$ $<$ $>$ $\neq$ $\cong$	(Generally found only in advanced mathematics.) R (The range of 'R' may be the entities named by: '$=$', '$<$', '$>$', '$\neq$', '$\sim$', '$\cong$', and '$\subseteq$'.)

ual constant which is a single symbol and which cannot be meaningfully broken down into smaller parts. Such atomic constants seem to be very few in number; there are the decimal digits '0', '1', and so on to '9' and then there are a few letters used as constants (almost always as abbreviations for compound individual constants); examples of the latter are 'i' (for '$\sqrt{-1}$') and 'e' (as a name for the base of natural logarithms). Any fraction is thus a compound individual constant, as is any numeral for a number greater than 9. It is especially interesting to us that we cannot find any individual constants for geometric objects in synthetic Euclidean geometry. (It is true that 'π' is an individual constant, but it names a number, not a point, line, or other geometric figure.) The letters 'A', 'B', 'C', and so on, are generally individual variables for points. If 'C' is used for circumference, it can best be analyzed as a function symbol; the same is true of many other words (and their related symbols) such as 'length', 'bisector', and 'area'. We thus have an example of what Russell probably meant in the quotation at the beginning of this section. Geometry is revealed as the study, not of particular geometric objects, but of relations between and functions of

The displayed sentences are essentially synonymous and in each case the correct reply is *2*. Emphasis on this synonymity is an important teaching device and should not be ignored. With this in mind, certain aspects of 9th grade algebra can be taught far earlier and to good advantage, as illustrated by the English writers Bass and Dowty who advocate this as early as the English equivalent of our 2nd grade.[2]

At this point, the phrase 'value of a variable' becomes especially important. Because of its frequent use, it is essential that it have an unambiguous meaning. This meaning should be simply 'a member of the range of the variable' or 'a referent of the variable'. There is nothing unusual about this, but we want to emphasize that a value of a variable is a thing and *not* a name of a thing. In the same way, we can extend the meaning of 'value' so that we can say that the value of a fractional numeral is a rational number and, even more generally, that the value of a numeral is its number.

Perhaps this is the time to focus our attention on the grammatical function of variable. This function is similar to that of indefinite pronouns and general nouns in everyday language. We shall make this clear by the use of an illustration. Consider the following sentence:

She said he was a member of it.

This sentence is neither true nor false, that is, it has no truth value. Only upon replacement of each 'she', 'he', and 'it' by a name of a person or an object will the sentence result in something which may be judged to be either true or false.

We shall call a sentence like the one above 'a propositional function'. This propositional function will yield a sentence capable of being judged either true or false only if *each* of the pronouns in the propositional function is replaced by a name of a unique person or object. Replacing only some of the pronouns by such names will result in other propositional functions. For example, each of the following is such a derived propositional function:

Nancy said he was a member of it

or

She said John was a member of it

or

Nancy said John was a member of it

Nancy said he was a member of Tallahassee Kiwanis Club.

To obtain a sentence from the propositional function, each of the pronouns in it must be replaced by a name of a person or an object. Thus,

Nancy said John was a member of Tallahassee Kiwanis Club

is a sentence which is either true or false.

Analogous to the concept of a range of a variable, the pronouns 'she' and 'he' have a range which is the set of human beings, and 'it' has a range consisting of organizations. For the sake of emphasis, we would like to point out again that the range in each case is a set of things and not names.

This leads us to seek analogous examples in the area of mathematics. Consider the following:

$$\text{For every } x \text{ and for every } y,\ x^2 - y^2 = (x - y)(x + y).$$

The intended meaning of this expression is that 'x' and 'y' in it can be replaced by names of numbers to obtain, in every case, a true sentence. Thus, '2' for 'x' and '3' for 'y' yields

$$2^2 - 3^2 = (2 - 3)(2 + 3)$$

which is a true sentence.

In conventional textbooks expressions like the one above are written without the use of quantifiers. They are commonly called 'identities'. Thus,

$$x^2 - y^2 = (x - y)(x + y)$$

is an identity because, no matter what numeral is put in place of 'x' and in place of 'y', '$x^2 - y^2$' and '$(x - y)(x + y)$' yield two names for the same number. We would have no serious objections to deleting the quantifier, provided the meaning of such expressions is made quite clear. This, unfortunately, is seldom achieved.

Let us consider another example:

$$\text{For some } x,\ x + 2 = \sqrt{3}.$$

Again, 'x' in this expression can be replaced by a name of any number. In this case, however, we will find that most of the sentences obtained will be false. For example, replacing 'x' by '$\sqrt{2}$', results in

$$\sqrt{2} + 2 = \sqrt{3}$$

which is a false sentence. We know that the only number whose name inserted in place of 'x' in '$x + 2 = \sqrt{3}$' will yield a true statement is $\sqrt{3} - 2$.

We need to pause here for some elaboration of the phrases, 'for every x and every y' and 'for some x'. The first phrase is usually called 'a universal quantifier' and the second 'an existential quantifier'. The universal quantifier 'for every x and every y' is frequently replaced by its synonym 'for all x and all y'. The existential quantifier 'for some x' has the synonyms: 'there exists an x such that', and 'there is at least one x such that'. Abbreviations are frequently employed for the quantifiers. 'For all x' is written as '$\forall x$', and 'for some x' as '$\exists x$'. So much for the notation. Now, what is the purpose of quantifiers and are they necessary?

The quantifier 'for every x and every y' serves as an abbreviation for the phrase: 'where 'x' and 'y' may be replaced by a numeral naming a number from a specified set of numbers'. In this sense, it serves a definite purpose—it places expressions, which may be otherwise meaningless and isolated from all context, in a meaningful setting.

At the beginning of this section, the advantages and economy gained through the use of variables were illustrated. The Distributive Principle of Multiplication over Addition is an additional illustration of this aspect of a variable. Let us state this principle as follows:

For every a, every b, and every c, $a(b + c) = ab + ac$.

In the first place, this expression covers a multitude of statements like:

$$2(3 + 7) = 2 \times 3 + 2 \times 7$$

$$-\tfrac{1}{2}(-\sqrt{3} + \tfrac{1}{3}) = -\tfrac{1}{2} \times (-\sqrt{3}) + (-\tfrac{1}{2}) \times \tfrac{1}{3}$$

$$(3 - 2i)[(-5 - \tfrac{1}{2}i) + (-7 + \sqrt{3}i)]$$
$$= (3 - 2i)(-5 - \tfrac{1}{2}i) + (3 - 2i)(-7 + \sqrt{3}i)$$

and so on.

Secondly, this expression conveys to us something which would be difficult and cumbersome to describe in the English language. Nevertheless, let us attempt to do it:

The product of a number and the sum of two numbers is equal to the sum of the products of the first number by the second number and of the first number by the third number.

After we have said it, we are still not sure whether we said what we wanted to say or whether anybody understands what we said or what we wanted to say.

Perhaps one more illustration bearing testimony to the advantages gained through the use of variables should prove of value. You recall that earlier in this chapter we agreed that a fractional numeral is a symbol (page 334). In order to describe what kind of symbol it is, we must resort to variables. We can do it as follows:

We call a symbol a 'fractional numeral' if it is of the form

$$\frac{*}{\Delta}$$

and can be obtained by replacing '$*$' and 'Δ' by names of integers, provided 'Δ' is not replaced by a name for the number zero.

Thus, '$*$' replaced by '-2' and 'Δ' replaced by '7' yields the fraction '$-2/7$'.

You will note that we have extended the notion of variables as letters to other symbols, like '$*$' and 'Δ' above. We believe that such practice should be encouraged in the classroom. Thus, in the examples we used in this section, it is immaterial what letters are used. For example,

For every x and every y, $x^2 - y^2 = (x - y)(x + y)$

could have been written as:

For every m and every s, $m^2 - s^2 = (m - s)(m + s)$

or as:

For every ○ and every ⬭, ○² − ⬭² = (○ − ⬭)(○ + ⬭).

Similarly, the Distributive Principle could be written as:

For every α, every β, and every γ, $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$

or as:

For every ○, every ⬭, and every Δ,

○(⬭ + Δ) = ○ × ⬭ + ○ × Δ.

In concluding the discussion of variable, we should remark that it is our belief that the view of the variable we attempted to convey in this section is quite adequate for the purposes of high school mathematics instruction. Teachers desiring to pursue this topic to a much deeper extent are encouraged to study references found in the notes at the end of this chapter.

DEFINITION

> ... A definition is no part of mathematics at all, and does *not make any statement concerning the entities* dealt with by mathematics, but is simply and solely a statement of a symbolic abbreviation: it is a proposition concerning symbols, not concerning what is symbolized. I do not mean, of course, to affirm that the word definition has no other meaning, but only that this is its *true mathematical* meaning.
>
> —BERTRAND RUSSELL

Perhaps the most appropriate way to begin this section is by defining 'definition'. Examination of a variety of written materials shows that there are three easily distinguishable uses of the word. There are two different processes called 'definition' and, third, the end product of either of the processes is also called 'a definition'. One process is that of examining the uses of a word or symbol and extracting from this a statement about the meaning of the word. This is *approximately* the process used in writing dictionaries. On the other hand, there is the process mentioned by Russell above of creating new words or symbols to serve as abbreviations for less convenient language or symbolism.

Despite the difference in the processes, they produce essentially the same result, namely, a statement that two words, phrases, sentences, or symbols, have the same meaning. This may be called 'a definition-statement'. When written in ordinary language, definition-statements may appear in a considerable variety of forms. When written completely in mathematical symbolism, they are considerably more uniform, although still having some variety. Even within this variety, three essential parts can always be discerned. Let us consider several examples.

(1) '2' is defined as '$1 + 1$'

(2) '$a - b$' means '$a + (-b)$'

(3) '$x \leq y$' will be used for '$x < y$ or $x = y$'

On the left of each of these statements is the piece of language being defined (technically known as the *definiendum*). On the right is the lan-

guage which does the defining (the *definiens*). (Note that right here are examples both of definition and of the value of definition. We introduce these Latin words not to be high-brow but because we anticipate a considerable amount of reference in the next pages to the language being defined and to the language that does the defining and we should like a more convenient way of doing this.) Between the *definiendum* and the *definiens* always appears some word, phrase, or symbol which expresses the idea that the *definiendum* and *definiens* have the same meaning. In our opinion, one of the best symbols for this purpose is that used by Whitehead and Russell, '= *df*'. It may be read, *equal by definition.* If the reader is of the opinion that there is some value in uniformity, then this symbol has much to recommend it. Thus the three definitions above would be written:

$$2 =_{df} 1 + 1$$

$$a - b =_{df} a + (-b)$$

$$x \leq y =_{df} x < y \text{ or } x = y.$$

There is considerable disagreement about several features of definition. One is the matter of whether we define words or things. The majority opinion among current mathematical logicians is in favor of defining words and symbols. Briefly their argument is that things have an existence independent of the names we may use for them and that all we can do is to decide how we shall use words and symbols (clearly, Humpty Dumpty, in our first quotation, was a mathematician). We concur in this judgment and also feel that it is pedagogically sound.

The second place for disagreement is in criteria for a good or proper definition. One position, which stems from Aristotle and his followers and is still followed in most high school geometry books, is in terms of genus and species, of no avoidable negative terms, and so on. The more modern position claims, among other things, that while this may be adequate for many or most individual terms, it is not applicable to operators or to relation symbols. To be specific, we do not believe that the genus-species matter can be applied to any of the three definitions above. The simplest and most important criterion of the modern position is that the *definiens* shall contain only words and symbols which have been previously defined or accepted as undefined.*

We mentioned earlier that there are two different processes which lead

* There are other much more sophisticated criteria established in modern logic. The interested reader is referred to Suppes' *Introduction to Logic.* These criteria are of little concern below the advanced university level.

nique is suggested here—namely, tn explain 'polynomial' as derived frora roots meaning 'many' and 'names' can be helpful tn the student. (It can also be the source of worse confusinn if not handled carefully, i.e., unless it is explained that the nriginal definition of the word is no longer used but that mathematicians have found it convenient to use 'polynomial' for expressions with as few as a single term.) That this is not explained in so many introductnry algebra books leads us to an interesting speculation: are tbese authors nverly influenced hy the notion that n word has a *real* meaning? Perhaps it was this kind of thing that led Carroll to have Humpty Dumpty ask: "Who is to be master, the word or I?" Taking an example at an earlier level, it seems botb interesting and helpful to young students to see that nur numeral '3' is probably derived from an earlier form with three hnrizontal tallies, thus '≡'. It is furtber valuable to show them the numerals for 2 and 3, used by modern Arabs, namely '٢' and '٣', to point out that this was probably derived from vertical tallies '||' and '|||'. An especially clear example of this is the creation of our sign '='; Robert Recorde's report of his cboice nf this as a sign for equality because he could think of no two things more equal than those lines certainly shnws that there was nothing random about his choice of a symbol.

Geometry is an especially good source of symbols of this general kind—for example, the use of '△' as an abbreviation for the word 'triangle' nnd the use of '⌒' as in '$\overset{\frown}{AB}$' tn indicate a circular arc joining A and B. To summarize, while mathematical words and symbols may be arbitrary in one sense, they are frequently suggestive nf their meanings.

An important teaching technique is a consequence of the above. It is our feeling that students also should be encouraged tn *create new words* and symbols if they are tn understand fully the nature of mathematical symbolism. For example, the ninth grader who suggested '≐' as a symbol for 'approximately equal' had made a great deal of progress. This is not a difficult thing for the teacher tn achieve, if he or she has the right attitude. Very young children are apparently fond of making up names (in sophisticated language, of making definitions). That older children do not do this so easily only indicates to us that they have learned some wrong things about words and symbols; in particular, that there is something sacred about the dictinnary, or even that God created names for things as well as the things themselves. Many teachers may feel wary about allowing students to create their own names and symbols, for fear that tbey will not learn the standard terminology and symbolism. This seems to us to be wrong on two grounds: (1) there is frequently no such thing as standard mathematical usage (witness 'billion' as meaning a

thousand million in the United States, but a million million in England and the several different notations for the decimal *separatrix* '2.5', '2·5' or '2,5'). (2) it is our impression that students who create their own symbolism, even if different from some standard symbolism, are actually better equipped to learn the symbols created by others.

EQUALITY

The central theme of this chapter, viz., the distinction between things and names, is of paramount importance in considering the concept of equality.

We begin with an example of a commonly used phrase stating that a quantity may be substituted for its equal. A moment's reflection leads us to the question: "How is it possible to have two equal quantities? Aren't we really talking about the same quantity?" Our answer to the last question is: "Yes, we *are* talking about only one thing!" Clarification of this statement is a goal of this section.

Following our favorite approach, we shall again seek examples in the English language. Let us examine the following statement:

Mamie = Mrs. Dwight Eisenhower.

What assertion is made by this statement? It is that 'Mamie' and 'Mrs. Dwight Eisenhower' are two names of the same person. No assertion is made that the two names are the same—it is very easy to see that one name is much shorter than the other.

What is true of things is seldom true of the names of these things. It is very easy to find examples where the opposite of what is true of things is true of the names of these objects. For example, the name 'Nile' is half as long as the name 'Okavango', but the river Nile is several times longer than the river Okavango.

Thus, a statement of equality does not assert that the names appearing on the two sides of the sign '=' are the same, but that the names refer to the same thing.

We all know that George Washington was the first president of the United States. The statement we just made asserts that the two names:

George Washington

and

the first president of the United States

are names of the same person. There will be no error committed if, in any statement, one name should be substituted for the other. Whether

we write 'George Washington' or we write 'the first president of the United States', we are referring to the same person.

What does what we have said so far have to do with mathematics and, in particular, with equality? The answer will be forthcoming immediately if we consider a simple example from arithmetic:

$$\tfrac{1}{2} = \tfrac{2}{4}.$$

To be consistent with what we have said previously in this chapter, '$\tfrac{1}{2} = \tfrac{2}{4}$' is interpreted to mean that '$\tfrac{1}{2}$' and '$\tfrac{2}{4}$' are two names of the same number. This number happens to be a rational number (page 335). There are many other things we can say about this number. For example, one of its names is the numeral '$\tfrac{75}{150}$'; another name is the numeral '$\tfrac{105}{210}$'; still another name is the decimal numeral '0.5', and so on.

The statement '$\tfrac{1}{2} = \tfrac{2}{4}$' does not assert that the names '$\tfrac{1}{2}$' and '$\tfrac{2}{4}$' are the same. Just as we decide by sight that the names 'Mamie' and 'Mrs. Dwight Eisenhower' are different, so do we see that '$\tfrac{1}{2}$' and '$\tfrac{2}{4}$' are different.

Furthermore, just as we may correctly say that Mamie is Mrs. Eisenhower, we may also say that the number $\tfrac{1}{2}$ is the same as the number $\tfrac{2}{4}$, or that $\tfrac{1}{2}$ is equal to $\tfrac{2}{4}$, or simply that $\tfrac{1}{2}$ is $\tfrac{2}{4}$. In each case we are speaking of the same entity: in the former case of the lady one of whose names is 'Mamie', in the latter case of the number one of whose names is '$\tfrac{1}{2}$'.

In order not to ignore the subject of geometry, we shall consider some examples commonly found in plane geometry textbooks. You are, no doubt, familiar with the usual definition of a line: *A line is the shortest distance between two points.* But, if a line is a distance, then it is a number indicating how many units of length there are in the line in question. In the same textbook in which a line is defined in this fashion, one is quite likely to read that *a straight line can be extended indefinitely in either direction*, meaning then that the *shortest distance can be extended indefinitely.*

Many difficulties stemming from an incorrect use of language, we feel, can be corrected with a slight amount of effort. The difficulty above, for example, can be avoided by taking 'straight line' as an undefined phrase. Then we can correctly say that the distance between two points in a plane is the shortest when it is measured along a straight line. We should point out that this statement is true in the Euclidean geometry, but it is not necessarily true in other geometries. For example, in a geometry in which the plane is the surface of the sphere, the shortest distance between two points is measured along a great circle.

The failure to distinguish the line from the distance has many unpleasant implications. For example, after defining a line segment as 'a portion of a straight line between two points', textbook writers use a common abbreviation for a line segment which is of the form 'AB'. Then they proceed to make statements such as:

$$AB \perp CD$$

and

$$AB = CD.$$

Obviously the first statement is about two line segments. From it we know that the line segment AB is perpendicular to the line segment CD. What about the second statement?

The second statement is not about line segments, for it would mean that 'AB' and 'CD' are names for the same line segment and that is obviously not what is intended by this statement. What is wanted is a statement to the effect that the measure of the line segment AB, say, in inches, is the same as the measure, in inches, of the line segment CD. We shall introduce the following notation to express this:

$$m^{in}(AB) = m^{in}(CD)$$

'$m^{in}(AB)$' is a name for the *number* of inches contained in the line segment AB. Hence,

$$AB \perp CD$$

means: *The line segment AB is perpendicular to the line segment CD.* And,

$$m^{in}(AB) = m^{in}(CD)$$

means: *The number of inches in the line segment AB is the same as the number of inches in the line segment CD.*

The symbolism '$m^{in}(AB)$', we must admit, is not very efficient; yet, we feel, it is worth sacrificing some efficiency for the sake of what we consider to be an understanding of a very important concept. After students achieve this understanding, it is quite appropriate to suggest that they seek better symbolism. Thus, they may prefer to use '$\overline{AB}$' in place of '$m^{in}(AB)$'. Of course, they will have to specify, for example, that '$\overline{AB}$' means that the line segment is measured in inches.

The difficulty created by the failure to distinguish between the *line segment* and the *measure of the line segment* finds its analogue in the use of names for angles. An angle is that which is formed by two rays emanat-

and

$$|x + 1| = -2$$

can be considered to be propositional functions. In exactly one instance the propositional function '$2x = 1$' yields a true sentence, whereas in no instance does '$|x + 1| = -2$' yield a true sentence.

So much for the concept of equation. What about the solution of equations? First, we should remark that it is sensible to talk about the solution of only those equations which contain at least one variable. It is out of place, for example, to talk about a solution of an equation like '$\frac{1}{2} = 0.79$'. We can quickly classify it as a false sentence and the case is dismissed.

Now let us turn our attention to the equation

$$2x = 1.$$

What does it mean to solve this equation? It means, to find all the numbers such that if any one of their names is put in place of 'x' in '$2x = 1$', a true sentence is obtained. In this case, this number is $\frac{1}{2}$, and we call it 'the root of the equation'. Note that a *root* is a *number*. Thus, the equation '$2x = 1$' has only one root, but that root, as any other number, has many names. It is immaterial which one of the many names for the number $\frac{1}{2}$ we choose to replace 'x' by—in every instance we obtain a true sentence. Thus,

$$2 \times \tfrac{2}{4} \;\; = 1$$

$$2 \times 0.5 \;\; = 1$$

$$2 \times 50\% = 1$$

are all examples of such true sentences.

It is our belief that such easy equations as the above should be used when first introducing the idea of equation. The students, after gaining an understanding of what it means to solve an equation, should be encouraged to use an intuitive way of solving these simple equations. For example, when encountering an equation like:

$$2x = 1$$

the student should reason as follows:

Two times what number equals one?

Aha, it's $\frac{1}{2}$! and so the equation '$2x = 1$' has the root $\frac{1}{2}$.

It is only proper to remark that we are aware of the fact that a great

many teachers make an excellent use of this intuitive approach to the solution of equations. *It is also appropriate to suggest that* more work can be done in the early elementary grades, preparatory to the formal treatment of equations, than is presently the practice. For example, it becomes quite appropriate to have young children handle questions like:

What number added to 5 gives 8?

and write it as:

$$? + 5 = 8$$

or

$$\square + 5 = 8$$

or

$$x + 5 = 8.$$

Each of the symbols '?', '$\square$', and 'x' *is a variable. Young children,* however, do not need to even see the word 'variable' in order to be able to properly understand the role it plays in this context.

Given an ample amount of practice in solving equations in the intuitive fashion discussed above, given a more complicated type of an equation, the student will soon discover that his intuition can carry him only a short way toward proficiency in solving equations. He will soon learn to desire more efficient methods. This is the time to suggest acceptance of the following properties of numbers (which may be labeled as 'Axioms' *in some developments):*

PROPERTY 1: *For every a, every b, and every c, $a = b$ if and only if* $a + c = b + c$.

PROPERTY 2: *For every a, every b, and every c, $a = b$ if and only if* $a \times c = b \times c$ $(c \neq 0)$.

Just a word of explanation of the phrase 'if and only if'. The statement of Property 1, for example, is actually two statements in one:

For every a, every b, and every c, if $a = b$, then

$$a + c = b + c$$

and

For every a, every b, and every c, if $a + c = b + c$, then $a = b$.

Once the student has accepted these properties, he is ready to handle more complicated equations. Let us solve, for example, the equation

$$2x + 7 = 3 - 3x.$$

Solution:

$2x + 7 = 3 - 3x$

$2x + 7 + (-7) = 3 - 3x + (-7)$ (by Property 1: '$2x + 7$' for 'a', '$3 - 3x$' for 'b', and '-7' for 'c')

$2x + [7 + (-7)] = -3x + [3 + (-7)]$ (by Commutative and Associative Principles for Addition)

$2x = -3x - 4$ (by arithmetic facts of signed numbers)

$2x + 3x = -3x - 4 + 3x$ (by Property 1: '$2x$' for 'a', '$-3x - 4$' for 'b', and '$3x$' for 'c')

$2x + 3x = -3x + 3x - 4$ (by Commutative Principle for Addition)

$5x = -4$ (by Distributive Principle)

$x = -\frac{4}{5}$ (by Property 2: '$5x$' for 'a', '-4' for 'b', and '$\frac{1}{5}$' for 'c').

It should be pointed out that in the step from '$-3x - 4 + 3x$' to '$-3x + 3x - 4$' the Commutative Principle for Addition is used as follows:

$$-3x - 4 + 3x = -3x + (-4) + 3x$$
$$= -3x + 3x + (-4) = -3x + 3x - 4.$$

In going from '$2x + 3x$' to '$5x$', the steps are as follows:

$2x + 3x = x \cdot 2 + x \cdot 3$ (by Commutative Principle for Multiplication)

$x \cdot 2 + 3 \cdot x = x(2 + 3)$ (by Distributive Principle)

$x(2 + 3) = x \cdot 5$ (by an arithmetic fact)

$x \cdot 5 = 5x$ (by Commutative Principle for Multiplication).

By placing the solutions of equations into such deductive framework, where what is done is based on basic principles and axioms, such meaningless expressions as 'transposition' and 'changing signs of numbers' are rendered unnecessary.

Closely related to the concept of equality is the concept of an identity. Consider, for example, the following which is ordinarily viewed as an

identity:

$$\sin^2 x + \cos^2 x = 1.$$

As it stands, the sentence is neither true nor false because of the variable 'x' and the way it occurs. What should have been written to make a true sentence is:

$$\text{For every } x, \sin^2 x + \cos^2 x = 1$$

which means:

For every replacement of 'x' by a numeral,

'$\sin^2 x + \cos^2 x = 1$' yields a true statement.

Thus, if 'x' is replaced by 'π', the sentence

$$\sin^2 \pi + \cos^2 \pi = 1$$

is true.

This and other examples lead us to the conclusion (really, an explication, in the language of the section on definition) that an identity is simply an equation with one or more variables, all of whose replacement instances are true. It is our claim that the kind of treatment described here can remove some of the difficulties students encounter in connection with equations and identities.

MULTI-MEANINGS OF WORDS

It is sometimes pointed out that mathematics harbors ambiguities because common words are used to describe extremely difficult mathematical ideas. Hence a *booby-prize* definition of mathematics: "Mathematics is the science which uses easy words for hard ideas."[4] This is a reversal of a situation which one finds in, for example, chemistry where such simple things as sugar, starch, or alcohol go by such unseemly looking names as "methylpropenylenedihydroxycinnamenylacrylic acid, or O-anhydrosulfaminobenzoine, or, protocatechuicaldehydemethylene."[4]

Thus, 'root' is used in the context of equations and has nothing to do with the picture of a thin elongated part of a tree buried in the ground which a youngster may envision when he first encounters the word. Here, a thoughtful teacher will exercise caution by developing the mathematical idea firmly before attempting to teach the child to associate a word with this idea. Furthermore, the sight of 'a square root' should not set one to thinking about equations because 'root' here is used to denote another mathematical idea. The preference for establish-

ing the idea before associating a word with it becomes still more mandatory.

'Product' should not send one dreaming about the many wonderful fruits of the manufacturer's labors, but should immediately bring on a thought of multiplication. But, beware, it isn't the multiplication of bacteria according to the rules of nature, but multiplication of numbers following the rules laid down by the mathematician. 'Function' should not bring on thoughts of pleasant social affairs or the processes in which a healthy gall bladder should engage. And 'table' used in connection with function should not make your mouth water at the imagining of a grand Thanksgiving dinner, rather it should turn your thoughts to a set of pairs of numbers.

'Power' is not intended to suggest the sway of some potentate over his subjects, or the dark of a room dispelled by the flick of a light switch. Rather it ought to turn one's attention to something like 'a^n' where 'a' goes by the name of 'base' and 'n' by the name of 'exponent'.

We could go on to the point of weariness in suggesting simple words which are used in mathematics to name quite abstract and rather difficult ideas. Perhaps the mathematicians should take a lesson from the medical man and the chemist and invent their own words for mathematical ideas to avoid danger of the associations already present in the case of many words.

Irvin H. Brune pointed to the danger of learning words without previous experiences to which to attach the words (21st yearbook, page 185). Thus, seemingly brilliant recitation of such phrases as 'Invert the divisor and multiply'; 'Cross-multiply'; 'Cancel'; 'Transpose'; 'Reduce'; 'Bring down'; 'Drop the per cent sign and move the decimal point two places to the left'; 'Annex the per cent sign and move the decimal point two places to the right'; 'Factor completely'; 'Double the width, double the length, and add'; 'Divide the number following *is* by the number following *of*'; 'Add the number of decimal places in the multiplicand to the number of decimal places in the multiplier', and 'subtract the number of decimal places in the divisor from the number of decimal places in the dividend, adding zeros to the dividend if necessary' may be just empty verbalisms. The user of these phrases may have no understanding whatsoever of the ideas and processes which he is able so generously and fluently to name.

In various places throughout this chapter, we have indicated the importance of understanding the distinction between things and names of things. In the discussion in this section it was pointed out that the student may be able to utter names of things without ever knowing any-

thing about the things themselves. It is one of the more important responsibilities of the teacher to acquaint the student with the nature of such things in mathematics as number, operation, function, triangle, circle, and so on.

Much of what was said in this chapter was suggestive of teaching which discourages students from uttering phrases which are devoid of meaning. The ability to repeat what one has read or heard the teacher say is an indication of good memory, but it does not offer a proof of the understanding of the ideas. Learning which is to last and lead to more learning must be based on an understanding of things.

See Chapter 11 for bibliographies and suggestions for the further study and use of the materials in this chapter.

NOTES

1. MENGER, KARL. *The Basic Concepts of Mathematics.* Chicago: The Bookstore, Illinois Institute of Technology, 1957 p. 7.
2. BASS, W. G., and DOWTY, O. S. *Counting and Arithmetic in the Infants School.* London: George G. Harrap & Company, Ltd., 1956. Chapters VI and X.
3. KASNER, EDWARD, and NEWMAN, JAMES R. "New Names for Old." *The World of Mathematics.* New York: Simon & Schuster, Inc., 1956. Vol. III. p. 1997.
4. *Ibid.*, p. 1996.

9

Mathematical Modes of Thought

E. H. C. HILDEBRANDT

> It is convenient to keep the old classification of mathematics as one of the sciences, but it is more just to call it a game.... Unlike the sciences, but like the art of music or the game of chess, mathematics is a free creation of the mind....
>
> It is an independent art, but, as it happens, it can be applied to the interpretation of nature,... Every new generalization gives a sense of power... One discovery suggests another; it does in fact *create* another.... Mathematics has a self-creating energy; the direction of advance is determined by the point that has been reached.—J. W. N. SULLIVAN*

THE IMPORTANCE OF DISCOVERY

THE GREATEST need in our present-day scientific age is for men and women who can use their minds as well as their knowledge of mathematics; for men and women who can use their understanding of the uses that have already been made of mathematics and apply it to new and unsolved problems in physics, biology, astronomy, the social sciences, and to new fields of technological knowledge still to be identified. Note that it is not merely necessary that the individual acquire a large amount of mathematical or scientific knowledge, but that the real test of his ability comes when he is confronted with a difficult problem-situation in science or further mathematics and is able to suggest or discover ways of finding an answer or the complete solution.

Understanding of mathematical concepts and some skill with its techniques are necessary to both the application of mathematics in new situations and to the creation of new mathematics, but these understandings and skills are far from sufficient. To apply and to invent mathematics one must also develop proficiency in *problem solving* or *reflective thinking.* To apply mathematics, and even more so to create new mathematics, one must not only be interested and curious, but also able and alert to

* J. W. N. SULLIVAN *The History of Mathematics in Europe, from the Fall of Greek Science to the Rise of Mathematical Rigour*, London: Oxford University Press, 1925, p. 7-10

perceive interrelationships between apparently different concepts, eager and able to see generalizations, analogies, special cases, and *idealizations*.

Just as there is no single set of rules for problem solving nor any infallible technique for creation so there is no neat prescription for teaching them. In general they must be taught implicitly and continuously by example and repeated exercises rather than explicitly by precept. Some suggestions for doing this are implicit both in the first part of Chapter 4 on *Inductive Reasoning* and in the discussions of *discovery* teaching techniques in Chapter 10. In the present chapter we wish to examine explicitly the mathematical modes of thought which lead to successful applications of mathematics and to the creation of new mathematics. We shall do this largely in terms of examples which we hope will simultaneously illustrate the principles we propose and suggest classroom procedures which will cultivate in students the ability to create and apply mathematics.

In this area of creation and application, individual differences will probably be even greater than elsewhere. This does not in any sense mean, however, that teaching designed to stimulate thought and develop problem-solving ability should be reserved only for the gifted. It does mean, however, that it is particularly important that the superior and gifted students be taught with these objectives in mind. For this reason we begin our discussion with comments on some characteristics of these children and we include in our examples problems and suggestions which are particularly appropriate for them at all grade levels. However, we would like to emphasize that with proper modifications in the amount of help given by the teacher, the size of the perceptive leaps expected of the students, and the time allotment given to the various steps of the process, *all students should repeatedly and continuously be "led" to discover or "invent" mathematical concepts and ideas for themselves. "Discovery"* classroom procedures can be used in developing standard classroom materials as well as in solving special problems such as those we treat in this chapter. Such discovery techniques help students to develop in their ability to think mathematically as well as in their understanding of the mathematics so developed.

Teachers of mathematics are faced with the question of how to identify students at an early age who will some day be able successfully to attack problems in the various sciences or branches of knowledge. The well-known intelligence quotient appears to be one means for identifying students with ability. Perhaps in the very near future, however, it will be found that there also exists some sort of science or problem-solving

A second type of problem is one which may require a certain amount of experimentation and assembling of pertinent data before convincing the student that a solution is possible. In case the solution is not unique, further consideration may lead to the need for acquiring new techniques and operations which have not heen studied previously. In his primary school days, the writer recalls that his mother received a game which was given away with the purchase of a pound of nationally advertised coffee (still being sold). The *gift* consisted of a small square cardboard box containing nine counters numbered from 1 through 9 with the directions that these should be placed in nine cells of a 3 by 3 unit square in such a way that the sum of every row, column, and diagonal of three counters should always equal 15. Every reader will recognize this as the simplest of magic squares. The game can be presented in such a way that it is a challenge to able children at any grade level. It does not requiro the discovery of a formula such as the one found by Gauss but does involve examining sets of three numbers, instead of two, whose sum is the same. Once the combinations which add up to 15 are listed, it will be found that there are only eight. Four of these contain the same number 5 which by coincidence fits the observation that the center cell of the 3 by 3 array requires a number which must be added in four different combinations. Examination of the other combinations soon indicates why the remaining odd numbers 1, 3, 7, and 9 cannot occupy the corner cells. This type of problem has merit in that it suggests the possibility of thinking about other sets of nine numbers which would add up to the same sum and whether these numbers bave to be consecutive or even form an arithmetic progression. A radio contest a few years ago made use of this modification when it required its participants to list as many different sets of magic squares as possible whose key sum for tbe rows, columns, and diagonals was 25 or less.

The third type of problem arises from situations which result from changing the conditions of a simpler one, for example, by adding another *dimension* to the problem, or requiring the study of a complete generalization or abstraction. For example, suppose in the lemon problem stated earlier we require the purchase of other fruit as well and state a problem sucb as this one:

> If lemons cost 3 cents and oranges 5 cents each, and we insist on spending all the money we have in our purse (which amounts to 8 cents or more) and on buying at least one of each item, are we always able to spend it all, and if so, what sets of combinations of fruit can we expect from our original wealth of n cents?

Similarly, instead of confining our attention to the 3 by 3 celled magic

are similar, semicircles, segments of ellipses and parabolas, or arches of sine curves and cycloids? Will the student who has heard about Fermat's last theorem be fooled when asked to replace Pythagoras' squares with cubes? Couldn't he also be stimulated to generalize the theorem to oblique triangles and discover the geometric form of the cosine law?

A fourth type of problem is the one which leads to the formulation of general principles, or to the conjecturing and eventual proof of specific theorems. For example, since the general quadratic can be solved by a formula which expresses both roots in terms of the coefficients, is it possible to secure formulas for all the roots of a general cubic and quartic? Once Cardan and Tartaglia had published their solutions to this question in the sixteenth century, it still took several more centuries before the question about formulas for the roots of the quintic and higher degree equations was found to have no answer. The famous Norwegian mathematician, Niels Henrik Abel, at the age of 22, made the discovery in 1824 that formulas for the solutions for the quintic and higher degree equations were not possible, not realizing that just a few years earlier, the Italian physician, P. Ruffini, had already found such a proof.

In this chapter, we plan to deal particularly with problems of the second and third type. Problems of the first type are found in many good texts, often at the end of chapters. They are usually more effective when they occur in general lists at the end of sections or in the appendix and, when properly used, become good teaching instruments. In the case of problems of type four, much more attention to the study of advanced mathematics is usually necessary before the student and mathematician achieve the understanding and mastery necessary for dealing with them.

It has already been noted that properly chosen problems not only are an effective means of identifying talented students, but that they may be instrumental in bringing out the important facets of the entire thinking process. Hence before describing some of the characteristics of other problems, it may be helpful if we give careful consideration to some of the stages of thinking that are present when we are attempting their solution. This may give us a better insight into ways in which thinking power in mathematics is developed.

THE PROBLEM-SOLVING PROCESS

The literature on problem solving and productive thinking uses a number of terms to describe some of the levels or stages of development such as *preparation*, *imagination*, *discovery*, *creativity*, *invention*, *orientation*, *incubation*, *contemplation*, *illumination*, *evolution*, *adaptation*, *substi-*

tution, *assimilation*, and *creative collaboration*. We will agree that some of these stages, as well as some of those listed below, are more prominent and important in *some problems* than in others. Also, there are certain levels which are not present in some cases at all. On the other hand, there are also some problems in which the later stages have never been reached. For the purposes of our discussion we will concentrate on eight levels which will be described briefly in the paragraphs which follow:

1. Presentation
2. Attention
3. Observation and Exploration
4. Classification
5. Further Exploration
6. Formulation
7. Generalization
8. Verification and Application.

Presentation. How the problem-situation is chosen and how it is presented to the students are factors which require the closest attention in preparing for the development of the thinking process. Unless the problem both has the inherent qualities indicating that significant conclusions or solutions are possible or plausible and it also becomes firmly fixed or registered in the mind, the successive stages in its solution will never be reached. There is a close parallel which we can draw between the process of problem solving and the growth of plants in nature. First of all, the mind (the soil) must be so prepared or conditioned that the problem (the seed) can be easily implanted (sown) in it. Just as the soil must have certain composition, texture, food content, moisture, and be deep enough for the roots to take hold, so the mind must be in that state of readiness where it is receptive to the challenge, i.e., there is a sense of curiosity and a willingness or desire on the part of the individual to deal with the situation. We also know that plants require sufficient sunshine and warmth of climate if they are to grow. In like manner we can say that the encouragement and interest of the teacher and the home in the child's mental development can do much to encourage proper use of his natural curiosity in dealing with a problem.

Suppose, for example, that we wished to introduce the notion of maximum and minimum as found in problems in the calculus to children in the intermediate grades. As an inducement to listening to the proposal, we might suggest special honors (a blue ribbon for the best answer might do) to the student(s) who constructed the largest open rectangular box which could be made from a piece of cardboard 5 by 8 inches by cutting a square from each corner and bending up the sides. If some reason for

producing such a box is required, we could add that we wished to fill it with some desirable substance (e.g., sugar or fudge) so that the box would contain the largest possible amount. The description of the problem must be presented so that a solution seems possible and at the same time we want the problem to register in the mind of the child. Also we are assuming that at the time the problem is stated, a certain readiness of the mind to cope with this problem has been attained, i.e., the child understands how to take measurements to eighths or tenths of an inch, he has mastered addition and multiplication of fractions and decimals, and knows how to apply the formula for the volume of a box when the dimensions are known. We also assume that he has had experience in cutting paper and in making boxes by folding paper or cardboard and taping or pasting the edges together. While any other convenient dimensions for the cardboard may be chosen, we have selected the present ones because the maximum value for the side of the square is a rational number. Also notice that sheets of paper for experimenting with this problem can be cut to size very easily by using standard sheets of 8½ by 11 inches. Teachers wishing to design other problems with rational dimensions can find several by experimenting with sets of three numbers a, b, and c which satisfy the Diophantine relation $a^2 - ab + b^2 = c^2$ where a and b are the sides of the original paper rectangle and the side of the cut-out square is given by the expression $(a + b - c)/6$. For example, in our case, $5^2 - 5 \cdot 8 + 8^2 = 49 = c^2$. Hence $c = 7$ and the side of the cut-out square would be $(5 + 8 - 7)/6 = 1$. The student, however, is unlikely to arrive at these formulas until much later, and perhaps not at all unless prodded and guided to extend his results. Let's return to his processes.

Attention. While the presentation level is one for arousing interest, the level which follows it should provide further stimulation to the child's natural curiosity. He may ask to have some of the parts of the problem restated. As his mind tries to digest the conditions of the problem, he may wonder out loud or to himself on the possibility of a solution. He may want to see a demonstration or may be willing to take scissors and a piece of paper and make a model of his own. The alert teacher will have some pieces of paper, 5 by 8 inches, ruler, and scissors available in order that every child may experiment with cutting out squares and folding up the sides of the box.

Seeds do not give evidence of producing shoots or sprouts in the first twenty-four hour period after they have been planted. In the same way, one should not be too concerned if students do not seem to respond the moment a new and somewhat difficult problem is proposed. Problems

require a period of time during which they can *germinate* in the subconscious or unconscious mind. Sometimes the teacher succeeds in hastening this period by raising certain questions which help to generate thinking. An attitude of inquiry and a flow of questions should always be encouraged.

Observation and Exploration. A single attempt at a solution is usually not enough. After several trials, it is possible to compare results. In this box problem, calculation will show that the contents differ as the height varies. Small squares ½ inch on a side cut out of the four corners of our paper will produce boxes which have large bases but are low in height. Squares 2 inches on a side will produce boxes with a smaller base but greater height. Some limitations on the dimensions will soon be evident, such as finding that the height can never exceed 2½ inches. Such results should be recorded so that possible changes can be observed or new ways for exploration can be suggested.

In using the term *observations*, we should note that we are not referring merely to what the eye sees, i.e., the actual cardboard box, but rather to a listing and examination of numerical and factual data which have been obtained from experiment and/or from calculations. The operation of recording these facts makes an impression on the mind which in turn may lead one to raise further questions or to make pertinent comments. Recorded facts can be compared with others. Euler stressed the importance of such observations when he wrote:

> It will seem not a little paradoxical to ascribe a great importance to observations even in that part of the mathematical sciences which is usually called Pure Mathematics, since the current opinion is that observations are restricted to physical objects that make impressions on the senses. As we must refer the numbers to the pure intellect alone, we can hardly understand how observations and quasi-experiments can be of use in investigating the nature of numbers. Yet, in fact, . . . the properties of the numbers known today have been mostly discovered by observation, and discovered long before their truth has been confirmed by rigid demonstrations. There are even many properties of the numbers . . . (in which) only observations have led us to their knowledge . . . in the theory of numbers, we can place our highest hopes in observations.

Even the secondary school student in mathematics should be made aware that there are some mathematical problems which have never gone beyond the *observation* stage in their solution. For example, C. Goldbach observed in 1742 that every even number he thought of was the sum of two prime numbers; for example, $2 = 1 + 1$, $4 = 3 + 1$, $10 = 7 + 3$, and so on. He conjectured from the observations that this was true of *all* even numbers. However, neither he nor any other mathe-

matician has been able to prove it or to find the single counterexample which would disprove it.

Classification. At this stage, we take stock of our progress. Some systematic way of organizing or tabulating our results may help. Sometimes a graph can be used to advantage. The individual who has acquired a desire to present his observations in an orderly and logical manner has found this stage to be an aid to a better understanding of the solution(s) of the problem.

In our box problem, we would list the length of the side of the cut-out squares in one column and in parallel ones show the corresponding length and width of the base of the box. Another column would show the calculated volumes. A study of the variation of the successive items might even reveal errors in calculation or in measurement which might not have been discovered otherwise.

In the 3 by 3 magic square problem, listing the combinations of three numbers whose sum was 15 was a help in identifying sets which could be used in columns, rows, and diagonals. In the original lemon problem, only one first degree equation in one unknown was really needed but the separate steps of simplifying this equation serve as a way of comparing equations so that the final solution can be readily recognized. The lemon and orange problem requires only one linear equation in two unknowns whose solutions are pairs of positive integers. However, when these solutions are listed in tabular form, they show several patterns of arithmetic progressions which might have escaped notice had not this form of classification been presented.

Further Exploration. A leveling off or a plateau in further progress may appear at this point. (This is found in nature's growth pattern, too.) Previously collected data may have to be re-examined. Other means for collecting facts may suggest themselves. Simpler ways of classifying results may be found. More encouragement from the teacher and the home may be needed while facts and solutions are re-examined. Perhaps some new patterns will come into focus. This may be the stage just prior to the one which Hadamard calls *illumination.*

It is here that we may wonder what will happen if our initial conditions are slightly modified or changed to see what resemblance may be found to solutions previously recorded. In the case of the box problem, we may wish to try a rectangular piece of material, 15 by 8 inches. If the sides of the cut-out squares are measured to tenths or to eighths of an inch, the dimensions of the *best* box would not be found among our recorded values, since the height would have to equal $1\frac{2}{3}$ inches. Such answers may be revealed only as our knowledge of the processes of algebra and

of the calculus is increased. In the case of the lemon and orange problem, we could change the prices to those listed currently in the markets and again find patterns which involve arithmetic progressions.

Formulation. This is the period during which various *hunches* or conjectures are suggested. Hence it might well he called the *hunch-period.* It is here that the mind begins to set up certain conditions and to draw some tentative conclusions which may result in theorems requiring proof. Some ohservations may suggest a single formula which will cover all cases. Possihly this is the stage that Polya has in mind when he says "*Let us teach guessing.*" On the other hand, at this point it may appear that the original prohlem is too hroad or involved and that it is advisable to examine certain suhcases or to modify some of the conditions. Then again it may he here that some extensions of the original problem seem possihle.

We have already referred to the particular Diophantine equation which arose in our box prohlem. The possibility of the existence of this formula can he expected from the special case we have used and further knowledge of maxima and minima as studied in the calculus. Hence a hy-product of this box problem is the study of this and similar equations.

Generalization. This level may appear in a number of different forms depending on the amount of experience and knowledge which the individual has acquired, hut teachers should always encourage students to look for generalizations and extensions of each result obtained or theorem studied. For example, when dealing with the relation $a^2 + b^2 = c^2$, one may desire those values for a, b, and c which are integers such as the sets 3, 4, 5 or 5, 12, 13. In that case, the ultimate generalization consists of the theorem that the set of three numbers, a, b, c, which satisfy the relations $a = m^2 - n^2$, $b = 2mn$, and $c = m^2 + n^2$, where m and n are relatively prime and $m > n$, will always have the property that $a^2 + b^2 = c^2$, and, further, all triples of integers having this property can he determined this way. However, some students may discover first that all numbers a, b, c, such that $a = m^2 - 1$, $b = 2m$, $c = m^2 + 1$, have the Pythagorean property. If so they should then ask themselves does this give all possible triples. Then as soon as they find a set of triples not of this simple type, the search for the ultimate formula will be on again.

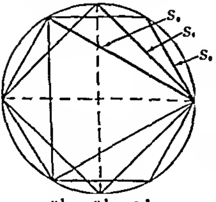
Suppose on the other hand we are considering the possibility that a and c are integers but b is not. Then we are led to the generalization that every odd prime can be expressed as the difference of two squares in one and only one way. A parallel generalization to this last case occurs

if we assume that a and b are integers and c is not. In this case, we reach the conclusion that a prime of the form $4n + 1$ can be represented as the sum of two squares. A generalization of the last statement but of higher order is the statement that a prime of the form $4n + 1$ is only once the hypotenuse of an integral-sided right triangle, its cube is three times, and so forth.

As soon as the student has become familiar with the fundamental trigonometric identities, he finds there are even irrational values that can be assigned to a and b in the relation $a^2 + b^2 = c^2$ and still produce a value for c which is integral; e.g., when a and b stand for the sine and cosine of the same angle, c is equal to 1. Here he even finds a generalization which had not occurred in Euclidean geometry; namely, that either a or b or both can be negative. Hence, important generalizations for all real numbers are discovered as one's mathematical knowledge is extended.

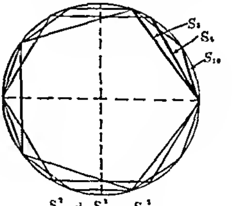
A geometric generalization of this theorem was suggested earlier when we noted it is still true if we replace the squares on the sides of the right triangle by regular polygons or by semicircles. Another substitution involves similar triangles on each of the sides, and that in turn raises the possibility of using similar polygons. Many other generalizations of this type can be found.

Pythagorean relationships may also occur between the elements of a set of geometric figures. The sides of regular polygons of n sides inscribed in the same circle provide an example. If, as represented in Figure 1, a and b are associated with the sides of a square, S_4, and of a hexagon, S_6, inscribed in the same circle, and c represents the side of the inscribed equilateral triangle, S_3, then $a^2 + b^2 = c^2$ again. Another set of related sides is shown in Figure 2. If S_5, S_6, S_{10} are the sides respectively of



$S_4^2 + S_6^2 = S_3^2$

FIG. 1



$S_{10}^2 + S_6^2 = S_5^2$

FIG. 2

the regular pentagon, hexagon, and decagon inscribed in the same circle, then $S_{10}^2 + S_6^2 = S_5^2$. A student noting these properties may wonder whether there are similar combinations of other regular polygons inscribed or circumscribed in the same circle.

The study of certain geometric curves may have its beginning in this same Pythagorean relationship. As an example we consider a right triangle with legs a_1 and b_1 and hypotenuse c_1, on whose hypotenuse we construct another right triangle whose leg $a_2 = c_1$ and such that the legs are in the same ratio as those of the original triangle. As this process is repeated (Fig. 3), each new vertex obtained can be seen to lie on some

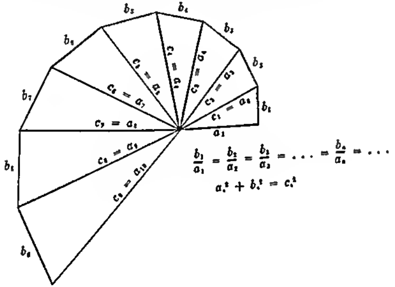


FIG. 3

curve which turns out to be the logarithmic spiral. The general equations for circles, $(x^2 + y^2 = r^2)$, ellipses $(b^2x^2 + a^2y^2 = a^2b^2)$ and hyperbolas $(b^2x^2 - a^2y^2 = a^2b^2)$ are Pythagorean in form. Intriguing to some are the curves resulting from replacing the exponent 2 in these formulas by ½, ⅔, and other values.

Verification and Application. A mathematical proof of every generalization is desirable. This proof may be either of the demonstrative or the inductive type. Some theorems are especially notable for the variety of proofs which have been discovered for them. The Pythagorean theorem probably holds the record for the largest number of different proofs.

One measure of the value of any problem-situation is the number of

uses or applications which can be found for properties which have been discovered. Its usefulness during various periods of history and in apparently unrelated areas is also a measure of its importance. The fact that the Pythagorean theorem is found on old Babylonian clay tablets of about 1800 B.C. and that at about the same period Egyptian surveyors used 3-4-5 triangles to lay out right angles indicate its usefulness from the very earliest days of recorded history. This same theorem is not only basic in such more recent and advanced mathematics as the analytic geometry and calculus and in situations arising in the various sciences, but we also find it applied in art where the whirling-square principle is developed in the theory of dynamic symmetry. One of the simplest dissection puzzles is one which requires that a set of geometric pieces be arranged in the form of a square and then reassembled to form two squares whose sides are not necessarily equal.

In the pages which follow, several additional problem-situations will be presented in detail to illustrate more clearly the successive steps that can take place in an exploratory study. Every teacher should realize that many of the topics usually treated in the elementary and secondary school courses in mathematics may be introduced and developed by the use of similar problems. One should remember, however, that the nature or choice of the problem and the manner in which it is presented are of the greatest importance. These factors also must be geared to the age level and experience of the group or individual.

HOW TO CUT A SANDWICH INTO HALVES

1. Presentation. The thought behind this problem is that the average person or well-known *Man-in-the-Street* takes it for granted that cutting a sandwich (made from slices of a square loaf of bread) into halves can be done in two and only two ways (Figs. 4*a* and 4*b*): either a cut made parallel to one edge of the sandwich thus producing rectangular halves or a cut across the diagonal resulting in halves in the shape of an isosceles right triangle. The first is probably the more popular of the two and requires the shortest line, but in both cases the halves of our *idealized* square sandwich are congruent. By *cut* we mean that line which separates the planar base of the sandwich into two equal parts. We are not concerned here with a three dimensional approach to carving the sandwich, especially if it should be of the triple decker or higher deck variety! This would be a natural and more difficult generalization of the problem we have chosen to work with.

Some aspects of this problem-situation may have been noticed by the child in preschool experiences. One mother, for example, reports that

she used to cut slices of toast for her children in various shapes and sizes and used these not only as an incentive for eating hut also for counting experiences and for developing acquaintances with various geometric shapes and their properties. At first, her ingenuity in obtaining variations in shapes and sizes was somewhat taxed until she discovered that

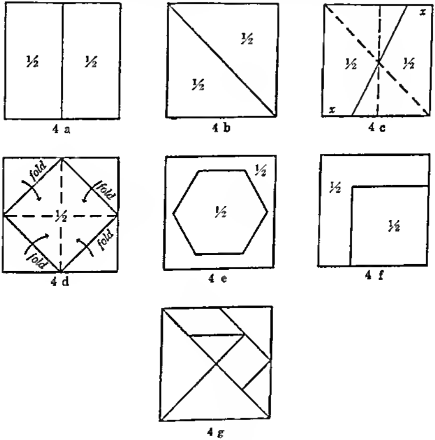


FIG. 4

her own children were quite capable of suggesting new *ways* to make such cuts also.

One appeal in this problem lies in the challenge to find a new way to perform a commonplace task, i.e., are there other ways than the two conventional ones for obtaining halves of a square? Connoisseurs of fancy sandwiches would have no difficulty in considering this approach and many a child likes the idea of *being different* when it comes to making or doing things.

The teacher might begin by sketching a square on paper or the blackboard. Three or four inch square pieces of paper should be placed in the

hands of every child with the suggestion that he attempt to fold the squares in the two ways indicated. Thus the *feel* of the straight-line cut is stressed and at the same time the learner may get a better understanding of the meaning of *halves.*

The problem part of the question can now be emphasized: Are there other straight-line cuts which will produce halves? If the child has had some experience with straight lines, then the question could be broached: Could these straight lines be modified so as to appear as broken lines and still produce halves? For an older child with an understanding of broken lines, the question can stress the possibility of using cuts which are not straight. In a high school class which has studied the areas of polygons and of closed curves, this presentation can pose the possibility that the halves need not be congruent.

Once the question is proposed, one should not expect a general rush to the kitchen with the intent of cutting up large supplies of bread slices for experimental purposes. It should be noted, however, that unless the presentation is so planned that the listener's mind is activated, i.e., begins to realize the possibility of the existence of new situations or solutions, the actual purpose here of getting the student to master various properties of geometric figures will not be attained.

2. Attention. In the earlier grades, it is important to have the child use concrete materials. Some children are capable of making sketches of squares which are quite satisfactory. If the child has had little experience in handling squares, it may be necessary to provide him with some three- or four-inch squares cut from squared paper which is ruled at half- or quarter-inch intervals. (These can be prepared easily from mimeograph stencils or from carbons used on other reproducing machines.) He can then count small squares and trace lines in such a way that he will have an equal number on each side of the trace or cut. He will soon notice a one-to-one correspondence between the small squares on both sides of this cut.

One of the first discoveries will probably be that if a line is drawn from points on opposite sides of the square and the same distance from diagonally opposite vertices, then this straight line will produce congruent trapezoids. (See Fig. 4c and note that our first two solutions were really special cases of this one.) Young children may have to cut the paper square apart in order to become convinced that the parts are congruent or equal. However, if such a result is not immediately forthcoming, some additional time may be needed and it may be well for the teacher to have extra paper squares available for practice purposes. The greatest benefit to be derived at this stage accrues when *every* member of the group is able to report his or her discovery at about the same time.

A comment on the value of experiences with paper-folding for children in the early grades is in order here. Japanese children, at an early age, for example, work with four- or five-inch paper squares cut from very thin paper, and the first fold made in most of their further projects is that of folding the square along a diagonal. Their ability to fold and refold parts of the square to form certain common figures (bird, monkey, box, balloon, elephant, and so on) requires the use of geometric figures throughout and the intricate patterns which result are real pieces of artful workmanship. While carrying on this paper-folding, called *origami*,[1] the Japanese child becomes well acquainted with many properties of geometric figures which are here being introduced and stressed in this problem-situation of the sandwich, but which he may study formally later in his course in plane geometry.

3. Observation. As soon as several different solutions are reported, the problem begins to reach this next stage. At first it may be noticed that there are quite a number of pairs of points on opposite sides of the square which are equidistant from the opposite vertices. This may be followed by the discovery that all of these lines, when shown in a given square, have one point in common, i.e., the center of the square. How long will it be before a student, in tearing or cutting the paper along this fold, will notice that if he digresses a small amount away from the fold in one direction he can restore the equality of the two areas by digressing an equal amount in the other direction from the fold?

Some children may feel that the half in the form of a rectangle is larger than the trapezoidal or the triangular one. They should be challenged to discover whether this conclusion is true. In seeking such verification, the child may find it when he places the triangular or trapezoidal half over the rectangular one and notices that parts of each overlap the other or are common to both and that the remaining parts are congruent.

Students in upper-grade classes may discover the *floating* figure principle which can produce halves. For example, when the corners of the original square are folded so that they form isosceles triangles and the original vertices meet at the center of the square as in Figure 4d, it will be found that a new square is formed which is just one-half of the original one. If we think of releasing this new square from its points of contact with the sides and allow it to *float* around inside the original square, then the remaining or uncovered part of the original square forms the second piece of the sandwich and can also be called one-half of the original square. This is a case of applying some take-away subtraction since $1 - \frac{1}{2} = \frac{1}{2}$.

Other *floating* figures can soon be imagined. Students with experience with regular polygons of more than four sides, can determine whether

these can be used as *floating* figures also (Fig. 4e). A real challenge to the understanding of any high school student with ability is the question whether this *floating* figure could be in the shape of an equilateral triangle. A test for their understanding of the formula for the area of a circle can be provided by requiring students to find the length of the radius of a *floating* circle—the hole in the sandwich—which equals one-half the area of the original square.

Another position for the *floating* square is to think of it as being shoved into one corner of the original square, thus having two of its sides and included angle in common (Fig. 4f). Children accustomed to using compasses will note the possibility of drawing an arc of a circle, using one vertex of the square as center, thus producing a quarter of a circle equal to half the original square.

Cases where certain suggested cuts cannot produce halves should also be observed. A good illustration is one in which we think of a single *bite* made into the sandwich starting from one side of the square and making this *bite* in the form of any triangle whose third vertex does not lie on the opposite side. Such a *bite* could not be a minor segment of a circle but variations of major segments are possible. Or, supposing that *bites* are made by parabolic *dentures*, one could practice with the area formula to determine the dimensions of a segment of a parabola which the fictional Paul Bunyan would have had to have cut out if he wished his first bite of a sandwich to have been one-half of it.

The challenge of these observations lies in the variety of situations which the imagination can produce. If the original squares are cut from cardboard of the same thickness, it should be possible to demonstrate with a set of beam balances that, no matter what their shape, the halves are always equal in weight. This of course will not furnish proof in the sense of mathematical logic, but it will furnish a verification of the students' conjectures and may at the same time be the stimulus for a valuable introductory discussion of the difference between proof and verification, deduction and induction, as discussed in Chapter 4. An interesting classroom display could be prepared by constructing mobiles in which the parts were all halves of the same square.

4. **Classification.** Some forms of classifying results have already been noticed as we have listed some of our observations. Our first thought was in terms of cuts which produced congruent figures with a straight-line cut through the center. However, other congruent figures were possible if digressions from such straight-line cuts were made in a symmetrical fashion using the center of the square as a point of symmetry. These cuts could also be included under those involving broken-line segments.

Another classification could be developed which involved the *floating*

figure principle. At first these figures were thought of as squares and regular polygons of more than four sides. The circle becomes the limit for such polygons. However, polygons which are not regular could also be considered and the possibility of using ellipses and limaçons need not be excluded.

The *bite* principle may be considered either from the corner of the square or as originating on one of the sides. These *bites* can be separated into those cases in which all lines are straight and those in which some lines are curved. The use of such curves as the parabola, hyperbola, catenary, cycloid, and sine curve could be proposed. Some ways of using an Archimedean or exponential spiral might also be suggested. A figure in the form of the snow-flake curve described by Kasner and Newman[2] would capture the imagination of some students.

5. Further Exploration. Part of this stage may require a certain amount of calculation. For example, a study of perimeters can be made based on the amount of original *crust* which each of the two parts possesses. We might impose the condition that the two parts have *crusts* which are in the ratio of 1:2 and study what limitations this places on the problem. The study of these ratios may be generalized to the form of $a:b$ later.

The need for the study of integral calculus may be noticed if we consider halves enclosed *partially* or *entirely* by some of these curves. How to approach the question of the areas of segments of such curves could become intriguing, too.

Part of this further exploration will result from *"what if you should"* questions. For example, what if we should cut the sandwich into thirds, fourths, or fifths, or cut the general square into any number of equal parts—would we be able to use the same principles regarding cuts?

Another direction our problem might take would be to substitute an equilateral triangle or other regular polygons for the square slice of bread, ultimately including the circle also. Various of the quadrilateral figures such as the rectangle, rhombus, parallelogram, and kite could be used. Some students will note an application of the *floating square in case the* original figure is a quadrilateral and the half is found by using the parallelogram formed by joining the midpoints of the sides successively.

6. Formulation. Some general principles that may be formulated were already suggested in the observation stage above. There we found that any straight line passing through the center of the square produces two congruent figures. Symmetrical digressions from this line also provided congruent figures. Equal area digressions or nonsymmetrical ones may produce equal figures also.

It has been quite evident that halves need not be congruent but are

always equal. Also straight lines need not be used for cuts but the location of curved lines requires careful calculation and will not produce the accuracy which rectilinear figures can attain.

7. Generalization. A succession of generalizations has appeared as this problem-situation was investigated. The conventional cuts of the *Man-in-the-Street* are merely special cases of the general one that any straight line drawn through the center of the square divides it into two equal parts. Certain digressions from these lines not only introduce properties which can be assigned to broken lines but indicate the opportunities which other mathematical curves present.

The *bite* and the *floating* figure principles lead to the best generalization of all: that any line—be it straight, broken, curved, or closed—could be thought of as moving across the plane to the square, then coming in contact with a vertex or moving across a side, and continuing across the square until it had reached a position where two distinct and equal parts of the square are evident.

8. Verification and Application. That two parts of a square of different shape but equal size can be found, required an understanding of area as enclosed by the lines already mentioned. Where the two parts are bounded by straight lines only, the knowledge of areas of various polygons as developed in plane geometry should be sufficient to verify the conclusions that have been suggested. In the case of areas partially or entirely bounded by other plane curves, the need for the calculus is soon noted.

Some application of properties found in this problem-situation is seen in certain figures of the jigsaw type. Here, there are opportunities for the child to become acquainted with fractional parts of a given square—fourths, eighths, sixteenths, and so on, as well as thirds, sixths, twelfths, fifths, and sevenths. He might be led to consider the case where a figure is separated into two parts which are in the ratio of $a:b$. Children in the early grades need such experiences with fractional parts of geometric figures including the circle. Actual handling of such concrete parts should prove valuable in preventing some of the absurd results relating to fractions which children often report in their practice exercises.

An extension or application of some of the properties of this square is found in the famous Chinese puzzle of seven pieces, the so-called tangram. This puzzle is sold in toy stores under various trade names (Fig. 49). The square (usually about 2½ inches on a side) is separated into seven pieces, two of which are right isosceles triangles each equal to one-fourth of the square; three more pieces are in the form of a square, an isosceles right triangle, and a parallelogram each of which equals

one-eighth of the square respectively; and the two remaining pieces are isosceles right triangles each equal to one-sixteenth of the square. Children in the earliest grades can make the parts either from a drawing on squared paper or by carefully folding a square along the lines shown in the figure. A good insight into the equality of figures is gained as two, three, or more of the pieces are tested to determine whether squares, trapezoids, parallelograms, or triangles can be formed from these pieces. H. E. Dudeney's *Canterbury Puzzles and Other Curious Problems*[3] contains illustrations of other figures which are not necessarily mathematical in design but all of which are equal in area. Perhaps some of these arrangements anticipated several works of modern art.

Another dissection which is used less frequently is the Loculus of Archimedes[4] shown in Figure 5. Archimedes may have aroused a mathe-

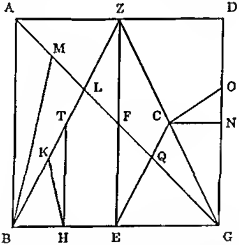


FIG. 5. The Loculus of Archimedes The square is divided into 14 parts with rational areas. Points Z, N, E, M, C, H are midpoints. TH is perpendicular to BG. K is the intersection of AH and BZ, O is the intersection of BC with DG.

matical interest among such royal friends as King Hiero of Syracuse or King Ptolemy Philadelphus of Alexandria by challenging them to entertain themselves and their guests by reassembling these fourteen pieces in various stimulating patterns. These pieces can be expressed as multiples of 48ths of the original square, i.e., they appear as $\frac{1}{48}$, $\frac{2}{48}$, $\frac{3}{48}$, $\frac{7}{48}$, and $\frac{8}{48}$. Once these fourteen pieces have been identified, many different arrangements of these pieces could be discovered which total $\frac{24}{48}$ of the original square. Making sketches for the cases where these parts have sides in common and where they could be thought of as being encased within a single border line could become another interesting diversion. At the secondary level, proof that these fourteen

pieces do actually bear the relationship to the whole indicated by the fractions listed above can test the mathematical ingenuity of the best of our present-day students of high school mathematics.

THE FOUR COLORS-FOUR NUMBERS PROBLEM

1. Presentation. This problem-situation arises from the question: Is it possible to take sixteen soldiers—call them $A1$, $A2$, $A3$, $A4$, $B1$, $B2$, $\cdots$, D3, and D4—representing four different companies (A, B, C, and D) and four different ranks in each company (1, 2, 3, and 4) and arrange them in a square formation of four men to a row in four rows in such a way that no two soldiers of the same company or rank are placed in the same row, column, or diagonal of the square? Another version might be: Can sixteen class or room officers from four different classes or rooms (freshman, sophomore, junior, and senior) each having four different officers (president, vice-president, secretary, and treasurer) be seated on an auditorium platform in a square arrangement consisting of four rows of four chairs each in such a way that no two classes or offices are represented in the same row, column, or diagonal? In this case these sixteen individuals could be designated by symbols such as P-9, V-9, S-9, T-9, P-10, V-10, $\cdots$, P-12, $\cdots$, T-12.

A colorful way of presenting this problem to a class or group is to use a set of sixteen square cards cut from cardboards of four different colors with the numbers 1 through 4 marked on separate cards of each color. (Stapling a piece of masking tape on the back of each card, with the *sticky* side exposed, makes it possible to press the cards against a blackboard or window pane and move them about as desired. Small plastic circular discs with adhesive on both sides, now available commercially, may also be used.)

Another way of presenting the problem is to use four sets of regular polygons of 3, 4, 5, and 6 sides respectively drawn on cards with numbers 1, 2, 3, and 4 appearing inside the respective figures. A teacher could prepare these figures on a mimeograph stencil and give the children an opportunity to arrange the cut-out pieces on their desk or work with them at home.

2. Attention. An "all thumbs" approach used by the teacher in trying to find a solution can easily arouse the attention of the children in the middle and upper grades. In the first demonstration, the teacher may place the cards or the symbols on the board so that some rows and columns seem to meet the requirements and give the impression that *it almost works*. The teacher may use chalk in four different colors which has the advantage that erasures can be made readily. Likewise the

children may choose colored pencils or crayons, say, blue, green, red, and yellow, and write the numbers in the sixteen cells of a 4 by 4 square. They will find this method cumbersome after a while, however, because changes are hard to make. When the students themselves suggest the possibility of using symbols in one color, they will be discovering a lesson on the importance of a wise choice of symbols as this relates to the development of mathematics. One notation which they may suggest is to use $B1$, $B2$, $B3$, and $B4$ to represent the four blue cards; $G1$, $G2$, $G3$, and $G4$ to represent the four green cards, and similar symbols for the cards in the other two colors.

It may be well to allow several days to elapse before going on. Students should be encouraged to hand in a copy of their solution as soon as possible and asked to indicate the approximate amount of time they spent in finding it. Some in the past have reported successes in less than fifteen minutes while others have persisted for two or three hours. At this stage, it is important to allow every person as much time outside of school as he cares to take so as to insure as many successes as possible rather than to hurry on to other aspects of the problem.

We are primarily concerned in observing ways in which learnings take place. Hence the teacher should lend encouragement but refrain from describing the solution herself. Showing a class or an individual *how* defeats the purpose and may result in loss of attention as well as interest. Time after time it has been noticed that those students who are shown answers either try to remember how the solution is obtained or lose interest in it entirely. This problem is not a difficult one, once a method for its solution has been discovered, and it provides elementary school children opportunities to develop their reasoning powers.

A trial and error approach to a problem has always been one of the first ways of investigating some new problem. Unfortunately it takes time and often leads to disappointments but frequently it is the first approach to the study of new problems. Persistence is one of the abilities often listed as an aim of mathematics education and this present problem may well prove to be a measure of the persistence of some individuals.

3. Observation and Exploration. Alert students may notice a parallel here with the requirements for the formation of a magic square. The sum of every row, column, and diagonal is always 10 with the added feature that colors are not to be duplicated. It is hoped that some students will begin to wonder whether this *color* feature can be translated into mathematical terms. If so, they are recognizing a parallel which arises in the art of programming for electronic calculators.

Whenever a systematic approach is reported, it is usually found that it suggests placing four cards of different numbers and colors in one of the rows, columns, or diagonals. In Figure 6*a*, cards $B1$, $G2$, $R3$, and $Y4$ have been placed in that order in the first row. Our next intent was to place $G1$ and $B2$ in the second row. Then the remaining cards for this row *had* to be $R4$ and $Y3$. Most persons report that after a certain point, the remaining cards just *fall into place*. This phenomenon suggests that there may be certain mathematical laws which apply here.

B1	G2	R3	Y4
R4	Y3	B2	G1
Y2	R1	G4	B3
G3	B4	Y1	R2

6*a*

Y2	R1	G4	B3
G3	B4	Y1	R2
B1	G2	R3	Y4
R4	Y3	B2	G1

6*b*

Y3	B2	G1	R4
R1	G4	B3	Y2
B4	Y1	R2	G3
G2	R3	Y4	B1

6*c*

A1	C4	D2	B3
D3	B2	A4	C1
B4	D1	C3	A2
C2	A3	B1	D4

6*d*

FIG. 6

Another observation which it is hoped will be made is that groups of four cards representing four different numbers and colors may be found in other positions besides in rows, columns, and diagonals. For example, in Figure 6*a*, the four corner squares satisfy this condition, as do also the four center squares. Elementary school children have a wonderful time noticing other such groups. The art of discovering such patterns is another one of the aims of mathematical education and is definitely one of the facets of the learning process.

4. Classification. After several solutions have been reported and examined, two questions may arise: (1) What constitutes a different solution? and (2) How many different solutions are there? In the case of the first question it will be agreed that if a given solution is rotated through an angle of 90° in either direction or through an angle of 180° that the resulting square is essentially the same as the original. In fact, one criterion for classification may be based on whether the same relative order appears except for a cyclical change in the elements. For example, if we take the first two rows in Figure 6*a* and rewrite them below the

third and fourth rows, we obtain another solution, since the diagonals are different (Fig. 6b). The order in the rows and columns is cyclically the same so this new square belongs to the same group of squares as the one in Figure 6a. Similarly if we think of beginning a solution by moving every element to the left and up one space starting with the element $Y3$ in Figure 6a, placing $R4$ at the end of the new first row, moving the original first row into the bottom row, and making similar cyclical replacements for the other elements for the fourth row and column, we obtain Figure 6c.

At this point it may also become apparent that comparisons may be made more readily if the cards are designated in some orderly or systematic fashion. An improvement in our use of symbols may make further study easier. Instead of using abbreviations for colors, it may be well to return to the letters used in the original designation of the soldiers in companies and use $A1, A2, \cdots, D4$, or something like 1-1, 1-2, 1-3, $\cdots$, 4-4. The former of these two seems to have some advantage at this stage. One should not ignore, however, the fact that the earlier use of colors has provided a worthwhile element of interest in this work. If our study of this problem had involved the fact that we wanted to know *how* to get a solution and then how to use this method to solve other problems like it, then we might have adopted some *best* notation at the very beginning. But if we believe in the importance of the successive stages of the learning process, then the repeated modifications provide us with experiences parallel to the various improvements and refinements which the mathematicians and scientists discover as they carry on the continuing phases of their investigations.

5. Further Exploration. One consequence of the preceding stage is the fact that new questions which need further investigation begin to take shape. Not only are we asking the question at this point, "How many different solutions are there to this question?" but we wonder about other questions such as, "Is it possible to determine the number of solutions by using some mathematical formula and without making a tabulation of all of the possible solutions?"

We are led also to the realization that certain phenomena need to be explored further. Such is the case with the observation that after a certain number of cards have been selected, the others seem to *just fall into place*. Suppose we select the four cards $A1$, $B2$, $C3$, and $D4$ and place them in the first diagonal in that order (Fig. 6d). If we select $A2$ as our fifth card, we find that it can be placed in only one of two positions, i.e., in the fourth cell of the third row or in the third cell of the fourth row. If we decide on the former, we find that the only card we

can choose for the second of these positions is $B1$, i.e., there were only two positions possible for the fifth card and only one for the sixth. In fact, once these six cards are in their correct position, the remaining ten find themselves assigned to unique places (Fig. 6*d*).

The possibility of determining whether there are other approaches to producing solutions needs to be explored also. One might consider placing the four cards $A1$, $B2$, $C3$, and $D4$ in the corner positions or in a block of four in one corner or in a center block and then determine whether the remaining cards occupy unique or semiunique positions. Or one might consider the famous *knight's move* used in magic squares of odd order. That there is some sort of *knight's move* principle present is noticed in Figure 6*d*. Using this movement, we can start from $A1$ and locate $A3$ and from the latter position on using the same movement locate both $A2$ and $A4$; likewise starting with $B2$ we can use this same movement to locate $B3$ and $B1$ and from the latter position obtain the place for $B4$. Similar locations follow from the $C3$ and $D4$ positions.

6. Formulation. The various stages involved in the development of solutions to problem-situations are not always clearly separated from one another nor is there necessarily a continuous path from one to the next. In our present exploration, we noticed that, in our 4 by 4 square, making a selection of a certain set of four elements and placing these in key positions was an important factor in the solution. Thus it may be possible for us to propose criteria for the selection of four elements and to indicate the conditions under which a solution may follow. Perhaps it will be possible to suggest also the various patterns for other solutions related to this one.

7. Generalization. Some students may wonder why the three colors-three numbers problem was not considered and they may have satisfied their curiosity on this score. On the other hand, a set of twenty-five cards for the five colors-five numbers case can be arranged in many solutions and may be studied in a manner similar to the one presented here. Unfortunately the six by six case cannot be solved—a fact discovered by the famous mathematician, Leonard Euler. Certain larger square arrays have been solved and their study can prove profitable for the able and ambitious student of mathematics.

8. Verification and Application. Proofs for some of the properties which have been indicated require mathematics beyond that generally studied at the secondary level.

The simplest application of the original solutions for the 4 by 4 square is to the development of magic squares of order 4. If the consecutive numbers in one color are simply written as the first four numbers, and

the same consecutive four numbers in another color are written as 5, 6, 7, 8, and so on for the other two colors, and these numhers are suhstituted in the solutions found, the resulting squares will he of the magic type. Not only will the sum of every row, column and diagonal he 34 hut other groupings of four squares, as already indicated, will provide patterns producing the same total.

Another interesting application is found in the theory of Graeco-Latin squares. An agricultural experiment may involve testing four different varieties of corn in comhination with four different soils or fertilizers but under tbe same climate or weather conditions. Not only can the separate comhinations be studied, hut the yield of a particular type of seed or of a certain soil may he shown to be superior under all conditions. Otber applications are found in statistics and topology.

PRIME NUMBERS

Introduction. One of tbe most fruitful fields for developing matbematical thinking is in tbe study of number theory. This is prohahly due to the fact tbat many questions ahout numbers require very little matbematical tbeory before one can appreciate the nature of tbe relationship involved. There is also a certain appeal to the imagination bere. Even little cbildren, after tbey have learned to count, like to talk about millions, quintillions, and trillions, not knowing wbich is larger but enjoying tbe appeal that bigness of numbers seems to have to tbem.

Some of tbe interest in tbe study of numbers is aroused wben one hegins to identify or list tbe so-called prime numhers. The recognition of tbose integers less than 100 whicb are evenly divisihle hy otber numhers as well as those remaining numhers which are not so divisible is absolutely necessary in the study of algehra.

As soon as children have learned those multiplication and division facts in which the product or dividend is less than 100, they should certainly he made more and more aware of the fact that there are several two-digit numhers which are not evenly divisible by 2, 3, 5, and 7. Soon they should he able to distinguish between odd and even numbers and to recognize that when any odd number is divided by 2, the remainder is always 1. By the end of the fifth grade, a child should be able to list all the primes less than 50 and hy the end of the sixth grade he should know all the primes less than 100 as well as the factors of all numbers in that range.

1. Presentation. For 2500 years or more, mathematicians have been trying to find simple ways of telling whether certain large numbers are prime or not. It is not difficult to tell when a numher is divisible by 2

or 5, and the usual test for divisibility by 3 or 9 should prove of interest to children after they have had some experience with division. But in the case of large numbers, it may be necessary to divide by several possible factors to test their primeness and often this work becomes tedious.

There is no reason why the method discovered by a Greek mathematician, known as the Eratosthenes sieve method, could not be introduced to children at fifth or sixth grade level. The fact that listing all the numbers from 2 to 100 and crossing out every second number after 2, every third after three, every number after 5 ending in 5, and every seventh number after 7 produces numbers which are not divisible by any other number less than themselves other than one, can be presented in such a way that children will always remember this method.

Certain questions for generating further thinking can then be presented or developed, depending on the age and grade level. Among these could be the following:

1. Do you believe that all prime numbers with the exception of 2 are odd?

2. Can the Eratosthenes method be extended for numbers between 100 and 200? Between 200 and 300? Can one use it between 500 and 600 without knowing all the primes less than 500?

3. In finding the primes less than 100 why did one not also cross out every 11th number after 11, every 13th after 13, and so on? In finding the primes less than 200, can one stop when one has reached the 7's step? Is there some way of knowing when one has done enough crossing out when finding the primes for any number less than a number one wishes to name, e.g., 2000?

4. Are there as many prime numbers between 100 and 200 as there are less than 100? If not, is it conceivable that the number of primes decreases for every successive 100? In that case, is it possible that eventually one might find an interval containing 100 numbers in which not a single number is prime?

5. Is it conceivable that if one continues examining larger and larger numbers that eventually one would find no more primes?

6. If one knows that a certain number is prime, is it possible to locate the very next prime after that?

7. What is the best way of determining whether some large odd number is a prime without constructing a complete sieve or consulting a table?

8. Do the gaps between primes increase in length? For example, the differences between primes below 75 are either 2, 4, or 6. Does that

mean that there will never be any with a difference of 8, 10, 12, or more? Is it possible that eventually one might no longer find a difference of 2 between successive primes?

Some additional preparation for these questions can be provided by giving students an opportunity to construct their own table of primes up to 500 or 1000 rather than providing such a table for them at this point. This becomes a good exercise in collecting data and at the same time provides some practice in accuracy. Later a table of primes less than 3000 should be provided and their attention called to the table of primes less than 10,000,000 of D. N. Lehmer.[5]

As soon as students have had some experience with formulas, several unsolved problems of prime numbers can be broached: Is there a formula which will always give a prime for every integral value of the variable? Does a formula exist or can one be found which will express *all* prime numbers?

2. Attention. By referring to a table of primes, it will be noticed that some of the numbers can be seen to be members of certain arithmetic sequences, e.g., 5, 11, 17, 23, 29, 35, 41, 47, 53, 59, $\cdots$. In fact out of the first 10 terms of this sequence, 9 are prime. The first 10 terms of the sequence based on the expression $6n + 1$ does not fare quite so well since our "batting average" is only 7 out of 10 and in the case of the sequences for $4n - 1$ and $4n + 1$ we get a *batting average* of .600. Questions of the following kind can be challenging: Does our batting average get better or worse as we consider the first 20 terms of either of these sequences? Can anyone discover a sequence in which successive terms have a common difference in which the first 10 terms are prime? If 100 per cent perfection in an expression is not possible, what expression does yield the largest number of primes for the first 50 or 100 values of the variable? A *contest* for determining some best method or formula does have an appeal value to many children.

When it begins to appear that 100 per cent perfection cannot be reached with arithmetic sequences, a slight variation of this same principle can be suggested, i.e., sequences in which the *differences* between successive terms are in arithmetic progression (actually this becomes a sequence whose terms satisfy a quadratic polynomial). Such an example is the sequence 5, 7, 11, 17, $\cdots$ or the one 11, 13, 17, 23, 31, $\cdots$ where the successive differences consist of the sequence of even numbers.

3. Observation and Exploration. It soon becomes apparent that none of the suggestions considered thus far will succeed in always providing primes. Attempts to create such sequences is good practice however. Various combinations for a and d in arithmetic sequences should